Physical and Genetic Containment of Transgenic Bacteria for Real World Applications

Physical and genetic containment of transgenic bacteria for real world applications

A dissertation presented by

Finn Stirling

To:

The Harvard Graduate school of Arts and Sciences

In partial fulfillment of the requirements
for the degree of:

Doctor of Philosophy

In the subject of:

Biological Sciences

Harvard University

Cambridge, Massachusetts

October 2019

Dissertation Advisor: Pamela Silver **Finn Stirling**

Physical and genetic containment of transgenic bacteria for real world applications

Abstract

Synthetic biology has promised and delivered on an impressive array of applications based on genetically modified microorganisms. While novel biotechnology offers undoubted benefits, it comes with an attached risk of both known and unknown negative consequences, particularly if transgenic organisms are allowed to proliferate freely. To achieve containment of an organism, confidence to a near scientific certainty must be shown that they cannot transfer their transgenic genes to other organisms and that they cannot survive to propagate in unintended environments. This thesis develops novel ways to achieve the required containment for the application of microbes in an uncontrolled environment. A recoding scheme is developed for altering codon assignment, creating a genetic firewall between an engineered and a wild-type organism that prevents horizontal gene transfer. Toxin-antitoxin based kill switches that respond to environmental signals such as temperature and pH allow for control over the location a transgenic microbe can survive in. These technological advances are put in the context of the wider field of containment, and what regulation would be required to consider a transgenic microbe safe to deploy is discussed.

Table of contents

Acknowledgements

This work was funded by DARPA [BRICS HR0011-15-C-0094]; as well as the NIH training grant [5T32GM007598].

I am truly grateful for the support I found in the Molecular and Cellular Organisms PhD program at Harvard. I would particularly like to thank Michael Lawrence, Patty Gonzalez and Fanuel Muindi, who never failed to answer a pestering email asking for help, direction or advice. The home department of the Silver lab, Harvard Systems Biology, has also offered all the support and welcome I could need. My time there has brought me many happy memories of karaoke, escape the room, and lighthouse shenanigans. Thank you to Kathy Buhl and Meghan van Orden and everybody else in sysbio operations who make everything run so smoothly, as well as to Scot Jandly for always being ready to show me how to restart my computer and for lending me my first D&D book. Special thanks go to Loden Dundutsang for the many, many hours he literally poured into making the 1000s of carefully buffered plates required to do all the pH sensitive assays in chapter 4. Without your care and dedication I would not be defending for another year!

II would like to thank all of the students who gave their time and energy to contribute to this thesis; Juliet, Rachel, Sherry, Sam, Elly, Hannah, Adam and especially Lizzy. I hope you learnt as much as I did.

I would like to thank my cohort of MChOmies for helping to make the last 5 years the best of my life. It has been a massive privilege going through this process with you all and I feel very lucky to have formed such close relationships with so many of you. I would also like to thank everyone who lived in or just hung around Brentwood over the years, and for the last few months those who made the move to Aldie, including all 8 cats and any wildlife they chose to bring home. Big shout out to Yossarian for just sort of existing, you are great at it. Special thanks need to go to Leo for being the go-to person to tell me

why my computer is making me sad by doing things I don't understand, and to Sean for his capacity to think like a snake to find missing things.

To the Silver lab I owe so much gratitude for making my grad school experience remarkably enjoyable and pain free. I don't remember any times when I was persistently unhappy, because if research stress ever got the better of us we had many remedies in the form of flunky ball, ski trips, D&D, camping trips, ice cream eating contests, volleyball, football, group meeting bingo and the enormous pile of board games I was allowed to sit next to. I would particularly like to thank Yu Heng, and James for collaborating on the genome recoding work, and Sasha for bringing together the counting systems with the kill switches. You were all a joy to work with. To Roger, my long suffering bay-mate I just want to say that somebody once told me the world is going to roll me, and I am sorry for not being the sharpest tool in the shed. To Shannon, for always showing me a stern face when I deserved it, and Marika, for being able to kick all the things, and to both for being a part of the extended bay family. To Cameron, who took me in when I first appeared in the Silver lab and showed me how exciting synthetic biology can be. And to everyone who ever contributed to baked goods Wednesday, made a cake for someone's birthday, brought delicious gifts back from trips overseas, gave a gift at secret Santa or just brought snacks to group meeting to keep us well fed. Thank you, you are both the reason the Silver lab has such a continuous strong community and why Jeff isn't allowed any more cookies.

Speaking of, a well-deserved special thanks must go to Pam and Jeff. Your guidance throughout graduate school has made me the scientist I am now and I am very grateful for all of the opportunities and freedoms you have allowed me to explore while carrying out my PhD. You are the shields that guard the realms of grad students.

Of course, I would not have achieved anything in life if it had not been for the constant love and support from my family. I want to thank my cousins and aunts and uncles for setting up a group chat when I left the UK so I would never feel apart from all their lives. I want to thank my family in New York

for offering me a home when I first visited the States, and for always opening your doors and brunch opportunities whenever I imposed myself. I want to thank all of my grandparents, Juliet, Peter, Charlie and Diana for the inspiration and academic support in following this path, for every trip to the park, every walk in the woods and every game of scrabble you let me win.

To my brother Alfie for being the one I always try just that little bit harder to keep up with, and my sister Freya, who will always be the one I look to for guidance to do the right thing. To my mum and dad, Topsy and Andy, you know words cannot begin to describe how much I owe you both for each and every day that you continue to show me your love and support, including the extensive editing of this thesis. I am so proud to be able to call you my parents, and when I raise a family I hope to be able to follow your example in how to offer the best possible existence in this world.

Finally, to my wife Sarabeth. It makes me so happy to look at each section of these acknowledgements and know you are a big part of every one of them. You are an essential gene to each aspect of my life, and I love you so much for it. You make me happier than I have ever been. You are my peace, moonlight, and the one who keeps me safe. My home.

Preface

"Absence of evidence is not evidence of absence"

Disclaimer

Chapters 2, 3 and 4 of this thesis were previously published in three separate journals, and are reproduced here without extensive editing. Inconsistencies in the style and formatting of the various sections are attributed to this.

Chapter 1 - Introduction

1.1 History of containment

Microorganisms have been used inadvertently in the processes behind food production for thousands of years, with various bacterial and fungal species playing key roles in brewing beer and the fermenting of wine, the baking of bread, and the process of turning milk to yoghurt. However, microbiology as a scientific discipline did not solidify until the 17th century, often predominantly attributed to the work of Anthony van Leeuwenoek in the 1670s. His passion for lenses and curiosity about the world around him led him to the first direct observation of bacteria as single celled organisms. The work of Louis Pasteur and Robert Koch, Fanny Hesse, Julius Petri, and many others over the next two centuries explored the possibilities of microorganisms as the vector for human and animal diseases. It was with these observations that the motivation to control and contain bacterial growth became apparent. Medical advances attributed to sterile and abiotic techniques played an integral part in the control of pathogens. In the 20th century, Alexander Fleming stumbled across the existence of antibiotics, leading to a revolution in bacterial containment by the application of small molecules.

Today we use microbes for an astonishing range of functions. In addition to the aforementioned production of ethanol, CO_2 and lactic acid for alcoholic drinks, bread and yoghurt respectively, both wild-type and engineered microbial species are used to produce a variety of industrially relevant products. Examples include; enzymes such as amylase (Sundarram & Murthy, 2014; Yoneda, 1980) and proteases (Razzaq et al., 2019); polysaccharides for use in food, cosmetics, and pharmaceutical products such as xanthan gum (Santos et al., 2000) and dextran (Sarwat et al., 2008); nutrients in the form of amino acids, nucleotides, vitamins and organic acids (Adrio & Demain, 2010); and pharmaceutical themselves, most infamously insulin (Baeshen et al., 2014), but also chemotherapeutics (Łukasiewicz & Fol, 2018) and antibiotics (Clardy et al., 2009).

In addition to their use as biological factories, the advent of synthetic biology has allowed microbes to be manipulated by humans to facilitate desirable processes, such as the bio-remediation of polluted environments and for use as living therapeutics. Bio-remediation involves the breakdown of an undesirable substrate, rather than concentrating on the production of a functional product. Applications include sewage treatment (Dhall et al., 2012), removal of pesticides (Paridah et al., 2016) and aromatic compounds (McClure & Venables, 1986) from soil, and clean-up of oil spills (Mapelli et al., 2017). Although microbes have been consciously used as therapeutics for over a century (Coley, 1893), the last decade has seen a plethora of possible applications presented. A non-exhaustive list includes diagnostic tools (Jonathan W Kotula et al., 2014; Riglar et al., 2017) and treatments for the human digestive system (Braat et al., 2006), a treatment for oral mucositis (Caluwaerts et al., 2010), HIV prevention (Lagenaur et al., 2011) and cancer immunotherapy (Zheng et al., 2017a). Various reviews have covered the novel applications that synthetic biology allows for bioremediation (de Lorenzo, 2009; Pieper & Reineke, 2000; Sayler & Ripp, 2000) and for living therapeutics (Riglar & Silver, 2018).

The European Commission has identified several negative consequences that uncontrolled proliferation of transgenic organisms produced from synthetic biology techniques could have on the environment. These include; uncontrolled propagation in the environment that may affect biological processes through interaction with indigenous populations, persistence in environmental niches that could lead to reduced biodiversity or the disruption of food webs, and the transfer of genetic information via horizontal gene transfer that could lead to adverse effects such as the spread of antibiotic resistance (Risks SCoEaNIH, 2015). Although the full risks associated with the spread of transgenic microbes are not known, it is this very uncertainty that makes it so important to implement tight control and regulation over microbial growth. The precautionary principle (Stirling, 2007) has established that the prime burden of responsibility falls on the proponents of a novel technology to

establish its safety to a high degree of scientific certainty rather than the reverse burden falling on others. This reflects the premise that "*absence of evidence is not evidence of absence*".

For these reasons, the field of biocontainment has been of interest for decades. The Asilomar conference (Berg et al., 1975) laid out guidelines for the introduction of auxotrophies to prevent escape of transgenic microbes. With the advent of synthetic biology, it has been judged necessary to repeatedly revisit the topic, with a number of reviews appearing in the intervening years (Chari & Church, 2017; Diwo & Budisa, 2019; Lee et al., 2018; Moe-Behrens et al., 2013; Molin et al., 1993; Schmidt & De Lorenzo, 2012; Torres et al., 2016; Wright et al., 2013). The applications of transgenic microbes covered so far can broadly be divided into two categories; applications where the entire intended purpose can be carried out in a controlled setting, and those where the intended purpose is designed to take place in an uncontrolled environment. Microbes engineered for production predominantly fall into the former category, whereas the majority of the second category are those microbes engineered to facilitate a process. However, both categories face the same requirements, that **i) that transgenic material is not spread to other species** and **ii) transgenic microbes are only viable in their intended environment**. Containment systems to date predominantly address one or both of these two issues.

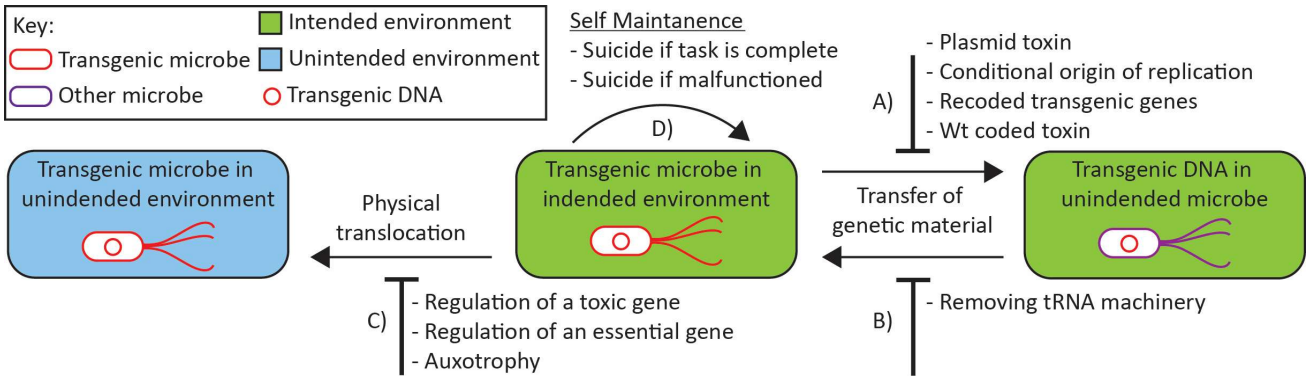

Figure 1: Schematic of biological containment mechanisms. A) Methods for addressing transfer and expression of genetic material in unintended hosts B) methods of preventing DNA from other bacteria being expressed in a transgenic strain C) methods for preventing the physical translocation of bacteria from their intended environment to an unintended environment, and D) methods for removing a bacteria from their intended environment if they are no longer required. Physical forms of containment such as barriers and waste management are not included.

A key means to quantify the effectiveness of a containment system used throughout this thesis, is to measure the 'survival ratio' it imparts on a population. **The survival ratio is the term used to define the ratio of colony forming units (cfu) in non-permissible conditions to the cfu in permissible conditions.** For a survival ratio of 10^{-4}, a population of 10,000 bacteria would yield only 1 survivor on transition to non-permissible conditions. Although not all systems can be quantified in such a manner (such as the essentializer in chapter 2), it is a useful tool for comparing different approaches. In the United States, the National Institute of Health has recommended a guideline providing that the "escape of the recombinant or synthetic nucleic acid molecule either via survival of the organisms or via transmission of the recombinant or synthetic nucleic acid molecule to other organisms should be less than 1 in 10^8 under specified conditions" (NIH FAQs, 2019, Apendix 1-1-B). A survival ratio of 10^{-8} would currently meet these requirements. For many applications, it is likely that this threshold is too high, which is discussed further in section 4.4 and 5.3.

1.2 Preventing the spread of transgenes

Every transgenic organism constitutes an example of one or more units of genetic material existing in an organism or in an arrangement that does not exist in nature. This enables interactions and transfers of genetic material that have not previously been possible. The various strata of life display both unique and interconnected mechanisms for spreading genetic material, and the intentional transfer of such material by humans is not only possible, but performed on a routine basis. The potential negative consequences of this can often be categorised as low risk, like the expression of fluorescent proteins such as GFP from the jellyfish *Aequorea victoria*. Other examples, including the spread of antibiotic resistance, are known to have dire direct consequences on human health. The use of GFP is considered harmless because no adverse consequences have been observed from its use in transgenic

strains (Lee et al., 2018; Moe-Behrens et al., 2013), and hypothetical problems that have been conceived of should transgenic strains expressing it be released are mild or uncertain. However this by no means indicates that if GFP were spread throughout an environment it did not originate from, that it is impossible for it to have a negative impact. Since it would be essentially impossible to prove the innocuousness of the spread of a transgene in all conceivable scenarios, a more feasible approach is to control its inability to spread in the first place with a reasonable level of confidence.

Early attempts at preventing horizontal transfer of genetic material between microbes relied upon toxin-antitoxin systems expressed on plasmids and genomes (Diaz et al., 1994; Torres et al., 2000; Diaz et al., 2003) or host repression of a plasmid borne toxin (Knudsen & Karlstrom, 1991) to prevent the spread of transgenic plasmids to other microbes. A similar approach has been proposed that uses origins of replication that are split between a plasmid and the host genome to prevent plasmid spreading (Wright et al., 2013). These to techniques were combined (Wright et al., 2015), achieving a survival ratio of $<10^{-3}$ for each method individually but the study did not report the survival ratio from the combined strain or use a low enough limit of detection to effectively ascertain the survival ratio of this approach. These systems are summarized in Table 1. It is of note that the stability (see section 1.4) of these systems, ie. the capacity of the system to maintain its function when passaged in permissible conditions, is not reported.

Year	Lead Author	Species	Conditional Origin	toxin/antitoxin system	Survival ratio	Stability
1991	SM. Knudsen	*E. coli*		*RelF*	10^-5	not tested
1994	E. Diaz	*E. coli*		colicin E3	10^-4	not tested
2000	B.Torres	*E. coli*		EcoRI	10^-4	not tested
2003	B.Torres	*E. coli*		*EcoRI, colicin E3*	10^-8	not tested
2014	O. Wright	*E. coli*	ColE2	*Kid* or *ζ*	<10-3	not tested

Table 1. Containment systems designed to prevent the spread of transgenic plasmids.

To achieve a 'hard lock' (where the probability of the transfer and expression of transgenic material can be deemed low enough as to be negligible) it is necessary to address the 'central dogma' of biology and unpick some of its core categories. Currently, all natural organisms are based around the

same genetic code. The four deoxyribose nucleotides are transcribed into the four ribose nucleotides, which are in turn translated using the ubiquitous triplet codon code into the twenty standard amino acids that make up all proteins. Several studies have sought to alter this dynamic in one way or another. By engineering a tRNA synthetase, it has been shown that a quadruplet codon scheme can be implemented in order to incorporate non-natural amino acids (Chatterjee et al., 2014; Hankore et al., 2019). In addition *E. coli* has been engineered to survive on an artificial nucleotide (Marlière et al., 2011) or to incorporate non-natural amino acid pairs (Malyshev et al., 2014).

Year	Author	Species	% recoded	Recoding method	Codons removed	Amino Acid
2011	F. Isaacs	*E. coli*	100	MAGE/CAGE	TAG	Stop
2016	N. Ostrov	*E. coli*	100[1]	Integrase based segments approach	AGA, AGG, AGC, AGT, TTA, TTG, TAG	Arg Ser Leu Stop
2017	YH. Lau	*S. typhimurium*	4.5	SIRCAS	TTA, TTG	Leu
2017	S. Richardson[2]	*S. cerevisiae*	100	SWaP-In	TAG	Stop
2019	J. Fredens	*E. coli*	100	REXER	TCG, TCA, TAG	Ser Stop

Table 2. Recoding approaches attempted to date. [1]this was achieved over 87 different strains. [2]The recoding of *S. cerevisiae* was a large collaborative effort that resulted in 7 publications in a special issue of Science (10th March 2017), of which S. Richardson is one lead author.

For containment applications, the most effective modification to the fundamental tenets of life upheld in the central dogma is to recode a genome so completely that it no longer requires the tRNA machinery for a specific codon or set of codons. This was first accomplished in bacteria with *E. coli* by removing all instances of the TAG stop codon and its corresponding tRNA (Lajoie et al., 2013), and again more recently removing TAG along with the serine encoding codons TCG and TCA (Fredens et al., 2019). Alternative methods have been published for the complete recoding of an organism such as an integrase based approach to achieve a 57 codon *E. coli* genome that is still ongoing (Ostrov et al., 2016), and an extensive international effort in *Saccharomyces cerevisiae* that successfully removed all instances of the TAG stop codon (Mitchell et al., 2017; Richardson et al., 2017). The different recoding approaches are

summarised in Table 2. A breakdown of each effort and the approaches used can be found in a review I co-authored with other members of the Silver lab (Kuo et al., 2018).

Freeing up a codon and subsequently its associated tRNA in this manner allows for 4 novel containment possibilities; 1) A recoded organism could have transgenic genes that are only translated in a functional manner by the host by inserting a recoded stop codon throughout the open reading frame thus initiating early termination of translation in any organism not recoded in the same manner. It should be noted this approach cannot prevent the transfer of transgenic DNA and subsequent mutations that remove recoded codons would allow expression (Figure 1A). 2) A more robust approach could be to incorporate a broad range highly lethal toxin, such as an endonuclease, alongside any transgenic circuit. Such a toxin could be prevented from expression in the transgenic strain by incorporating canonical codon usage. Upon transfer to a wild-type strain, toxin expression would cause cell death. As transformation and transduction of genomic material rely upon homologous recombination, transgenic circuits would reliably be transferred in single units (Figure 1A). 3) A recoded organism without the full contingent of tRNA's is unable to translate most novel genetic material it acquires, blocking a potential avenue of evolution. (Figure 1B). 4) Novel tRNA synthetases that incorporate a non-natural amino acid can be used alongside an unassigned codon to engineer an auxotrophy, forming an effective containment system (Figure 1C) (Mandell et al., 2015; Rovner et al., 2015), explored further in section 1.3 (Figure 1C).

A major obstacle to the widespread adoption of genome recoding is the ability to devise recoding schemes for a diverse range of organisms. Recoding efforts so far have concentrated on *E. coli* and *S. cerevisiae*, two of the most comprehensively studied model organisms. Recoding a genome requires an extensive knowledge of an organism's essential and overlapping genes, as well as established protocols for engineering and growth. In chapter 2, I report the development of the recoding technique Step-wise Integration of Rolling Circle Amplified Segments (SIRCAS) in the context of

recoding the genome of *Salmonella typhimurium.* Although less frequently studied, *S. typhimurium* shares many characteristics with *E. coli* with an 84% sequence similarity (McClelland, 2000). This study therefore simultaneously introduced a novel method for genome recoding and displayed the capacity to recode a relatively poorly characterised model organism.

1.3 Containment of transgenic microbes to their intended environment

All functional, currently employed applications transgenic microbes are undertaken in a controlled environment. Here, a controlled environment is defined as one where physical barriers, effective waste management and enforced regulation can easily be assured, including both laboratory research and industrial production. There are many published containment systems that have been designed to prevent transgenic microbes form escaping their intended environment. These generally fall into one or more of the following categories: allowing for the expression of an essential gene, regulation of a toxic gene, and supplementation of an auxotrophy (Figure 1C). In addition, regulation of a toxic gene has been shown to modulate the self-maintenance of a transgenic strain (Figure 1D). Table 3 compiles a reasonably exhaustive list of 35 containment systems, that take advantage of essential and toxic gene regulation, engineered auxotrophy or a combination of several mechanisms. It includes their mode(s) of containment, the signal(s) they respond to, their survival ratio, the evolutionary stability of the system when grown in permissible conditions and whether they are designed for use in a controlled or uncontrolled environment. Each row of the table represents a single containment system in a single strain. Papers that published multiple systems have generally been represented with whichever system had the lowest survival ratio, unless the modes of regulation are so disparate as to be considered separately of interest. Unless otherwise stated, the following systems are solely applicable to the processes described in Figure 1C.

Containment systems for a controlled environment

Of the 35 systems, 22 respond to one or more small molecules intended to be provided by human intervention, and are therefore designed to contain the transgenic strains to a controlled environment. Twelve systems consist of a basic design concentrating on a single molecule regulating a single toxic gene (Balan & Schenberg, 2005; Bej et al., 1992; Bej et al., 1988; Knudsen & Karlstrom, 1991; S. Knudsen et al., 1995; Kristoffersen et al., 2000; Li & Wu, 2009; Molin et al., 1987; Recorbet et al., 1993) or one or more essential genes (Agmon et al., 2017; Cai et al., 2015; Kong et al., 2008).

The last decade has seen a progression towards more complex containment systems. Three examples from the Collins lab (Callura et al., 2010; Chan et al., 2015) explore the capabilities of synthetic circuits to respond to multiple inputs controlling the expression of both toxic and essential genes, and are excellent examples of modular and customisable containment systems. Another impactful example comes from the Isaacs lab (Gallagher et al., 2015), who combined an auxotrophy, suppression of two essential genes and the expression of a toxin to construct a containment system with a survival ratio below their detection limit of 10^{-12}.

A unique containment system based on cell density (Huang et al., 2016) depends on collective expression of an antibiotic resistance gene whose expression is stimulated by the quorum sensing factor N-acyl homoserine lactone (AHL). At high cell densities, enough antibiotic resistance is expressed at a population level and released (upon cell lysis, expressed intracellularly, or excreted) to allow individual cells to survive. However if individual cells split off from the “microbial swarmbot”, insufficient resistance proteins would be expressed to allow survival. Although this system currently requires the continued application of antibiotics to provide the toxic function, it is suggested that future iterations could modulate the expression of toxic or essential genes, removing the requirement for supplementation and facilitating its use in an uncontrolled environment.

The final 5 systems designed to be used in a controlled environment are all auxotrophies, categorized as such because they require the uptake of the small molecule they are responding to for some aspect of their metabolism that cannot otherwise be self-synthesised. To construct a system that responded to atmospheric CO_2 (Clark et al., 2018) the CO_2 concentrating mechanism of the cyanobacteria *Synechococcus sp.* PCC7002 was removed. This only allowed survival when the transgenic strain was grown in an environment with at least 5% ambient CO_2, with atmospheric levels of CO_2 resulting in a survival ratio of 10^{-9}. Another study engineered an *E. coli* phosphite auxotroph by removing all phosphate production pathways except via phosphite uptake, achieving a survival ratio below their detection limit of 10^{-12} (Hirota et al., 2017). A system for making Synthetic auxotroph's based on a Ligand-Dependent Essential genes (SLiDE) was developed (Lopez & Anderson, 2015). SLiDE was used to develop a strain were the function of three essential genes was dependent upon the presence of the ligand benzothiazole. Finally, two synthetic auxotrophs were designed using the *E. coli* recoded strain with all instances of the TAG stop codon removed mentioned in section 1.2 (Lajoie, Rovner, et al., 2013). Novel tRNA machinery was introduced to incorporate non-natural amino acids into the primary protein sequence of several essential genes (Mandell et al., 2015; Rovner et al., 2015). These last two methods both showed extremely robust containment, below the detection limit of 10^{-11}.

Containment systems for an uncontrolled environment

The other 13 of the 35 systems in Table 1 were designed for application outside of a controlled environment. A thymidine auxotrophy containment system for *Lactococcus lactis* engineered to act as a therapeutic for Crohn's disease when applied to the human digestive system showed a survival ratio of 10^{-7} both *in vitro* and when applied to a porcine model (L Steidler et al., 2003). Although unable to propagate outside of a controlled environment, the functional application for this strain of expressing

human interleukin-10 did not require DNA replication, allowing its application in an uncontrolled environment. This system is particularly of note as being the only containment system to be tested in a clinical trial with humans (Braat et al., 2006), although with inconclusive success.

Five systems were designed for application in microbes engineered to facilitate the bioremediation of soil by the degradation of benzoates. While the degradation target, benzoates, are present, toxin expression is repressed allowing survival. Because of this they respond to the presence of small molecule regulators, but will initiate population suicide upon the completion of their task or translocation from their intended location without the need for human intervention (Figure 1C and 1D).

A system was developed that was designed to terminate a bacterial population upon a loss of function mutation, based on the bacteriophage lambda *cI*/*cro* regulatory system (Stirling et al., 2017). In the presence of either cI or Cro, the toxin is repressed. However, in the absence of either, repression is relieved and the toxin is expressed. This system allows for the maintenance of functionality of a specific transgenic strain, checking to see if the transgenic element is present and terminating the strain if it is not (Figure 1D), and is explored further in chapter 3.

Three systems that responded exclusively to temperature have been published, each with a different mode of control. The first controls expression of the *Serratia marcescens* nuclease *nucA* with the temperature sensitive mutant of cI, cI^{857} (Ahrenholtz et al., 1994). At 42 ^{o}C, cI^{857} is unable to function, and the toxin is expressed resulting in a survival ratio of 10^{-5}. The second controls expression of the DNA gyrase inhibitor *ccdB* using a mutant *tlpA* repressor that is activated below 36 ^{o}C. Using this system they were able to show the containment of a bacterial population to the mammalian gut with a survival ratio of 10^{-5} (Piraner et al., 2016). The third uses the regulatory region from cold shock protein A to control expression of the toxin *ccdB*, also achieving containment to the mouse gut with a survival ratio of a little over 10^{-5} (Stirling et al., 2017), and is reported in greater detail in chapter 3.

Finally, three systems that respond to pH are reported in chapter 4 of this thesis (Stirling et al., 2019). The first controls the expression of the toxin Doc with the pH sensitive promoter P_{asr}, achieving a survival ratio of 10^{-6} when exposed to pH 5 conditions. The second combines the pH sensitive expression of Doc with the temperature sensitive expression of CcdB to achieve a survival ratio of below 10^{-11} when grown at 22 ^{o}C and at pH 5. The third and final containment system uses an excisionase based system to only express *doc* upon two, non-consecutive exposures to low pH.

Year	Lead Author	Species	Auxo-trophy	Essential gene	Toxic gene	Responds to	Survival ratio	Stability	Used in
1987	S. Molin	*E. coli*			*hok*	tryptophan	10^{-4}	not tested	con
1988	AK. Bej	*E. coli*			*hok*	IPTG	$<10^{-6}$	not tested	con
1991	A. Contreras	*E. coli*			*gef*	benzoates	10^{-6}	not tested	un
1991	SM. Knudsen	*E. coli*			*relF*	IPTG	10^{-8}	not tested	con
1992	AK Bej	*P. putida*			gef	IPTG	10^{-5}	not tested	con
1993	G. Recorbet	*E. coli*			*sacB*	sucrose	10^{-3}	not tested	con
1994	I. Ahren-holtz	*E. coli*			*nucA*	temp. (cI857)	10^{-5}	not tested	un
1995	SM. Knudsen	*E. coli*			*relE*	IPTG	10^{-7}	not tested	con
1996	MT. Munthali	*P. putida*			*col*E3	3-methyl benzoate	N/A	not tested	un
1997	P. Szafranski	*P. putida*			*strept-avidin*	3-methyl benzoate	10^{-7}	not tested	un
1998	L. Molina	*P. putida*			*gef*	3-methyl benzoate	10^{-8}	not tested	un
2000	P. Kristo-ffersen	*S. cerevisiae*			*relE*	galactose	N/A	not tested	con
2001	MC. Ronchel	*P. putida*			*gef*	3-methyl benzoate	$<10^{-9}$	not tested	un
2003	L. Steidler	*L. lactis*	thym-idine			thymidine	10^{-7}	not tested	un
2005	A. Balan	*S. cerevisiae*			*nucA*	glucose	10^{-5}	not tested	con
2008	W. Kong	*S. typhimurium*		*asdA, murA*		arabinose	10^{-4}	not tested	con
2009	Q. Li	*E. coli*			*nucA*	arabinose	N/A	not tested	con
2010	JM. Callura	*E. coli*			*ccdB/* λ lysis*/ lexA3*	aTc, arabinose, IPTG	10^{-3}	not tested	con
2015	CTY. Chan	*E. coli*		*murC*	*ecoRI*	aTc, IPTG	$<10^{-5}$	Unstable after 4 days	con

2015	CTY. Chan	*E. coli*		*murC*	*ecoR1*	IPTG, gal, cellobiose	$<10^{-8}$	Unstable after 4 days,	con
2015	RR. Gallagher	*E. coli*	Biotin	*ribA, glmS*	*ecoRI*	aTc, IPTG	$<10^{-12}$	stable for >110 gen.	con
2015	Y. Cai	*S. cerevisiae*		HHTS, HHFS		galactose, estradiol	$<10^{-10}$	not tested	con
2015	DJ. Mandel	*E. coli*	bipA			bipA	$<10^{-11}$	not tested	con
2015	AJ. Rovner	*E. coli*	pAcF α			pAcF α, arabinose	$<10^{-11}$	not tested	con
2015	G. Lopez	*E. coli*	benzo-thazole			benzo-thiazole	10^{-7}	not tested	con
2016	DI. piraner	*E. coli*			*ccdB*	temp.	10^{-5}	not tested	un
2016	S. Huang	*E. coli*		*blaM*, *cat*[2]		cell density	10^{-5}	not tested	con
2016	R. Hirota	*E. coli*	Phos-phite			phosphite	$<10^{-12}$	not tested	con
2017	N. Agmon	*S. cerevisiae*		SEC17		estradiol	10^{-8}	not tested	con
2017	F. Stirling[1]	*E. coli*			*ccdB*	temp.	10^{-5}	Stable for >140 gen.	un
2017	F. Stirling[1]	*E. coli*			*ccdB*	cI and/or Cro	N/A	Stable for >140 gen.	un
2018	RL. Clark	*S. sp.* PCC7002	CO_2			CO_2	10^{-9}	not tested	con
2019	F. Stirling[1]	*E. coli*			*doc*	pH	10^{-6}	Stable for >100 gen.	un
2019	F. Stirling[1]	*E. coli*			*ccdB/ doc*	temp., pH	$<10^{-11}$	Stable for >100 gen.	un
2019	F. Stirling[1]	*E. coli*			*doc*	pH[3]	10^{-4}	not tested	un

Table 3. Containment systems designed to prevent escape of transgenic microbes. Strains: *Escherichia coli*, *Pseudomonas putida*, *Saccharomyces cerevisiae*, *Lactococcus lactis*, *Salmonella typhimurium, Synechococcus sp.* PCC7002. [1]explored in this thesis. [2]These genes are essential in the presence of the antibiotics carbenicillin and chloramphenicol. Each row represents a single containment system in a single strain. [3]This containment system only expresses the toxin upon two, non-consecutive exposures to low pH.

1.4 Evolutionary stability

Containment systems are only effective if they maintain functionality throughout the entire period of their intended use. Microbes exist across all domains of life, but are unified in their almost universal short generational lifespan and capacity to mutate and evolve at a rapid rate. This presents a problem for technology designed to contain and control the growth of engineered microbial species. To coin a term from the fictional Dr Ian Malcolm, "*Life, uh, finds a way*". Any containment system that is designed to prevent microbial growth inherently has the potential to inflict a fitness defect, and therefore an evolutionary pressure to remove this fitness defect. For different forms of containment, this can occur in different manners.

For an auxotroph based containment system, an engineered strain would require the capacity to produce the metabolite that it is auxotrophic for, or negate the necessity for it in the first place, which can be achieved in 4 ways. 1) By taking on genes or operons from other strains that confer this capacity, through horizontal gene transfer or sexual reproduction, most likely from related species. 2) Evolution of pathways that are already intrinsic to the engineered strain that become capable of producing the required metabolite. 3) Evolution or modifications to the composition of the other organisms in the surrounding environment that increase production and/or excretion of the metabolites in question, allowing an environment that was initially non-permissible to become viable. 4) In the case of the systems mentioned that rely upon essential genes requiring ligand cofactors or nnAA for function/ translation (Lopez & Anderson, 2015; Mandell et al., 2015; Rovner et al., 2015), SNPs and small mutations to the genes in question can result in removing those requirements.

Escaping from a system of containment that uses the regulation of an essential gene can come about through 3 potential mechanisms. 1) Mutation of the regulation of the gene in question, making it no longer dependent upon whatever factor was designed to control expression. 2) Evolution of other

pathways within the cell to fulfil the role the essential gene confers, negating the essential genes necessity. 3) Taking on the same or similar genes from other organisms that are regulated in an alternative manner.

Population control based upon the expression of a toxic gene has 4 possible modes of escape. 1) Mutation of the ORF of the toxin, negating its function 2) Mutation to the regulation of the toxin, either in its promoter or any transcriptional factors or in the regulation of the antitoxin (if present, see below) 3) Mutation in the target of the toxin. 4) Uptake of a resistance gene through horizontal gene transfer or sexual reproduction. Option 1, and in many cases option 2, represents a loss of function mutation, meaning the range of possible mutations that achieve the effect of escape is vastly wider than that required to confer a gain of function, which all other modes of escape mentioned above are some form of. Another factor that contributes to the instability of containment system based on regulation of a toxic gene is the inherent leakiness of most regulatory systems. Promoters can best be described as up or down regulated rather than off or on, as even when repressed or not induced, low levels of expression can be observed. To counteract the fitness defect imparted from leaked expression of a toxic gene in permissible conditions, an antitoxin can be included in the design of the system. Most natural toxins evolved to function alongside their cognate antitoxin, and this dynamic can be exploited by expressing antitoxin at low levels to negate the effect of a leaked toxin. This technique is explored in more detail in chapters 3 and 4.

Of the 35 containment systems reported in Table 1, only 7 provide data of their long-term stability when passaged in permissible conditions. Five of those are containment systems reported in chapters 3 and 4 of this thesis, and as stated all make use of toxins and respective antitoxins. The multi-layered containment system (Gallagher et al., 2015), consisting of a biotin auxotrophy as well as the arabinose mediated regulation of the two essential genes *ribA* and *glmS* and the toxin *ecoRI* nuclease,

was shown to be stable after passaging in permissible conditions for at least 110 generations. This system includes the *ecoRI* methyltransferase, the antitoxin to *ecoRI* nuclease. The Deadman and Passcode containment systems (Chan et al., 2015) also regulate the expression of *ecoRI*, as well as the essential gene *murC*, this time without the presence of *ecoRI* methyltransferase. Neither maintained their respective level of containment after a four-day period of growth, with an increase in survival ratio of 4-5 orders of magnitude. Passcode was additionally passaged for 4 days in *E. coli* MDS42pdu Δ*recA* (Csörgo et al., 2012), a strain lacking recombinogenic and mobile genomic elements. This reduced the Passcode escapee rate by 3–5 logs over the four-day period. This practice is appropriate for certain purposes, although the fitness defect observed in *E.coli* MDS42pdu Δ*recA* will prevent its widespread application.

1.5 Containment standard practices

Currently publications on containment systems are disparate in their reporting techniques. Not all reports display a quantifiable survival ratio by comparing cfu at permissible conditions to cfu at non-permissible conditions, and instead rely on metrics such as a growth curve to show growth disparities. Although it is easy to infer an effect from a growth curve, the fact that OD measurements do not differentiate well between living and dead cells means that a true survival ratio is hard to calculate. I recommend the survival ratios for containment systems should always be calculated by plating dilutions of culture in both permissible and non-permissible conditions and comparing cfu. It should be noted that for some systems an accurate survival ratio is fundamentally difficult to ascertain, such as the essentializer circuit reported in chapter 3. This system relies upon the efficiency of a P1 transduction to ascertain the number of escapees, therefore having a limit of detection for a single experiment is a survival ratio of between 10^2 and 10^3.

In most cases however, individual studies report unique limits to the sensitivity of their survival assays. This limit predominantly comes from the total cfu that are given the possibility to grow in non-permissible conditions. Using the traditional techniques of spreading culture on agar, it appears from both literature (Gallagher et al., 2015; Mandell et al., 2015; Rovner et al., 2015) and my own work reported in chapter 4 (F. Stirling et al., 2019) that the upper limit of plating a bacterial culture is reached at about 10^{11}-10^{12} cfu. Beyond this, the concentration of cells is so high as to cause clumping that prevents accurate assays. Increasing the area of agar plates used circumvents this, but quickly becomes prohibitive in the area of plates required. I recommend that when reporting the survival ratio for a containment system, it should become standard practice to reach this limit of detection. Lower survival ratios could feasibly be detected using technology such as a morbidostat (Toprak et al., 2013). Although this capacity is currently not easily accessible for all labs, widespread adoption of this technique may become necessary as the field develops.

The majority of studies did not report data on the stability of their respective containment systems. For all containment systems, no matter the application, evolutionary stability is an essential quality, and can easily be assayed by passaging in permissible conditions whilst calculating the number of generations that pass. A comparison of survival ratios before and after this growth period allows for a simple display of evolutionary stability. I recommend that determining the stability of a strain over 100 generations should become standard practice when reporting a new containment system. This is sufficient to allow at least one adaptive sweep to pass through the population, allowing beneficial mutations to take over (Maddamsetti et al., 2015, Novick & Szilard, 1950) (Sections 3.2, 3.3 and 3.4). Additional experiments comparing the growth rates of engineered strains with parent strains, either individually or in competitive co-culture growth assays, would also be informative. Further theory on the

capacity of asexual strains to maintain a circuit that confers only a very slight fitness disadvantage can be found in the discussion of chapter 3 (Stirling et al., 2017).

Chapter 2 - Large scale recoding of a bacterial genome by iterative recombineering of synthetic DNA

Yu Heng Lau[1,2], Finn Stirling[1,2], James Kuo[1,2], Michiel A. P. Karrenbelt[1,3], Yujia A. Chan[1,2], Adam Riesselman[4], Connor A. Horton[1,2], Elena Schafer[¨1,2], David Lips[1,2], Matthew T. Weinstock[5], Daniel G. Gibson[5,6], Jeffrey C. Way[1,2] and Pamela A. Silver[1,2,*]

[1]Wyss Institute for Biologically Inspired Engineering, Harvard University, 3 Blackfan Circle, 5th Floor, Boston, MA 02115, USA,

[2]Department of Systems Biology, Harvard Medical School, 200 Longwood Avenue, Alpert 536, Boston, MA 02115, USA,

[3]Systems and Synthetic Biology, Wageningen University, Dreijenplein 10, 6703 HB Wageningen, The Netherlands,

[4]Program in Biomedical Informatics, Harvard Medical School, Boston, MA 02115, USA,

[5]Synthetic Genomics, Inc., 11149 North Torrey Pines Road, La Jolla, CA 92037, USA

[6]Synthetic Biology and Bioenergy Group, J. Craig Venter Institute, 4120 Capricon Lane, La Jolla, CA 92037

* To whom correspondence should be addressed. Tel: +1 617 432 6401; Email: pamela silver@hms.harvard.edu

This article was first published in: Nucleic Acids Research, Vol. 45, No. 11 6971–6980, 2017
doi: 10.1093/nar/gkx415

My contribution

Although the overall project was formulated before I joined the Silver lab by Jeffrey Way and Pamela Silver, I played a key part in shaping the direction it was taken in, and the methods used as the project progressed. The computational design of the recoded organism was initially carried out by Michiel Karrenbelt, but I was involved in quality control, observing and correcting several errors in design that would have delayed the project by months. I took part in negotiating, communicating and ordering from the commercial DNA synthesis companies. I constructed the recombinant strain of *Salmonella typhimurium*, and alongside Yu Heng Lau developed the protocols required to transform large (>10 kb sized fragments) of DNA using the lambda *red* recombinase system, a technique integral to the success of the project. I helped develop and implement the protocols for homologous assembly of yeast artificial chromosomes, as well as their propogation by rolling circle amplification. When it came to writing the final manuscript I was involved with original text, editing and figure creation.

2.1 Abstract

The ability to rewrite large stretches of genomic DNA enables the creation of new organisms with customized functions. However, few methods currently exist for accumulating such widespread genomic changes in a single organism. In this study, we demonstrate a rapid approach for rewriting bacterial genomes with modified synthetic DNA. We recode 200 kb of the *Salmonella typhimurium* LT2 genome through a process we term SIRCAS (stepwise integration of rolling circle amplified segments), towards constructing an attenuated and genetically isolated bacterial chassis. The SIRCAS process involves direct iterative recombineering of 10-25 kb synthetic DNA constructs which are assembled in yeast and amplified by rolling circle amplification. Using SIRCAS, we create a *Salmonella* with 1557 synonymous leucine codon replacements across 176 genes, the largest number of cumulative recoding changes in a single bacterial strain to date. We demonstrate reproducibility over sixteen two-day cycles of integration and parallelization for hierarchical construction of a synthetic genome by conjugation. The resulting recoded strain grows at a similar rate to the wild-type strain and does not exhibit any major growth defects. This work is the first instance of synthetic bacterial recoding beyond the *E. coli* genome, and reveals that *Salmonella* is remarkably amenable to genome-scale modification.

2.2 Introduction

The next widely anticipated breakthrough in genetic engineering is the ability to rapidly rewrite the genomes of industrially relevant microbes, plants, and animals. Rewriting entire genomes will deepen our understanding of the genetic code and dramatically transform human health, food and energy production, and our environment (Annaluru et al., 2014; Boeke et al., 2016; Hutchison et al., 2016; Lajoie et al., 2013; Ostrov et al., 2016; Richardson et al., 2017). A major challenge identified by the Genome Project-Write consortium is the efficiency of building and testing large modified genomes (Boeke et al., 2016). In particular, genome recoding involves synonymous replacement of all instances of specific codons throughout an entire genome (Lajoie, Rovner, et al., 2013), requiring efficient assembly of large constructs containing thousands of designed base changes (Ostrov et al., 2016). New foundational technologies are therefore crucial for accelerating the pace of genome synthesis and modification.

New organisms based on recoded genomes can serve a variety of useful functions (Boeke et al., 2016; Lajoie et al., 2013; Ostrov et al., 2016; Wang et al., 2016). While recoded genomes still encode for the same translation products, the replaced target codons can be repurposed for incorporating non-standard amino acids into proteins and peptides (Chin, 2012; Kipper et al., 2017; Zheng et al., 2016). Loss of the target codon tRNAs in a recoded organism can also impair its ability to express foreign DNA acquired by horizontal gene transfer or viral infection (Lajoie et al., 2013; Natalie Jing Ma & Isaacs, 2016). This may enhance the genetic containment of an organism, a desirable feature when considering safety and stability of engineered organisms for applications in open environments (Moe-Behrens et al., 2013).

Several recent examples of modifying bacterial genomes illustrate the challenges associated with the construction process. Church and co-workers reported a piecewise strategy for debugging their design of a 57-codon *E. coli* genome (Ostrov et al., 2016). Testing 55 different strains containing independent ~50 kb segments of their recoded genome design on plasmids, 44 segments conferred viability. Although the

remaining 11 segments contained lethal elements, the piecewise strategy helped to expedite the troubleshooting process by narrowing down the number of possible offending mutations, enabling a design flaw to be fixed in one of the segments. While cryptic ribosome binding sites and changes in mRNA folding have been identified as potentially important factors when recoding (Kelsic et al., 2016; Napolitano et al., 2016), underlying principles of genome redesign are generally still not well-defined and fixing design flaws remains challenging (Lajoie et al., 2013). The empirical nature of recoding is exemplified in recent work by Chin and co-workers, in which recoding of an essential 20 kb region of *E. coli* was attempted (Wang et al., 2016). Three of the eight different codon replacement schemes gave viable organisms and one lethal scheme could be rescued by fixing a single offending codon, while the remaining four schemes led to inviability. Furthermore, lessons from the *de novo* synthesis of a 'minimal bacterial genome' by Venter and co-workers have implications for genome recoding (Hutchison et al., 2016). While the authors halved the genome size of *Mycoplasma mycoides* by removing non-essential elements, this required iterative testing to address complexities such as quasi-essential genes and synthetic lethal gene pairs. Similar issues may arise when implementing genome-wide recoding.

Herein, we report a streamlined method for accumulating large-scale recoding in a single strain of *Salmonella enterica* serovar Typhimurium LT2 (*S. typhimurium*), bypassing the reliance of previous methods on site-specific enzymes. *S. typhimurium* is of interest in engineering the gut microbiome for diagnosis, vaccination and treatment of human disease (Frahm et al., 2015; Kong et al., 2012; Zheng et al., 2017). An attenuated and recoded *S. typhimurium* could serve as a stable chassis for housing various diagnostic and therapeutic circuits that can autonomously detect and act upon changes in the gut environment (Kotula et al., 2014). In this work, we accumulate 1557 codon changes in 176 genes across 200 kb of the *S. typhimurium* genome (Figure 1), using an iterative process we term "stepwise integration of rolling circle amplified segments" (SIRCAS).

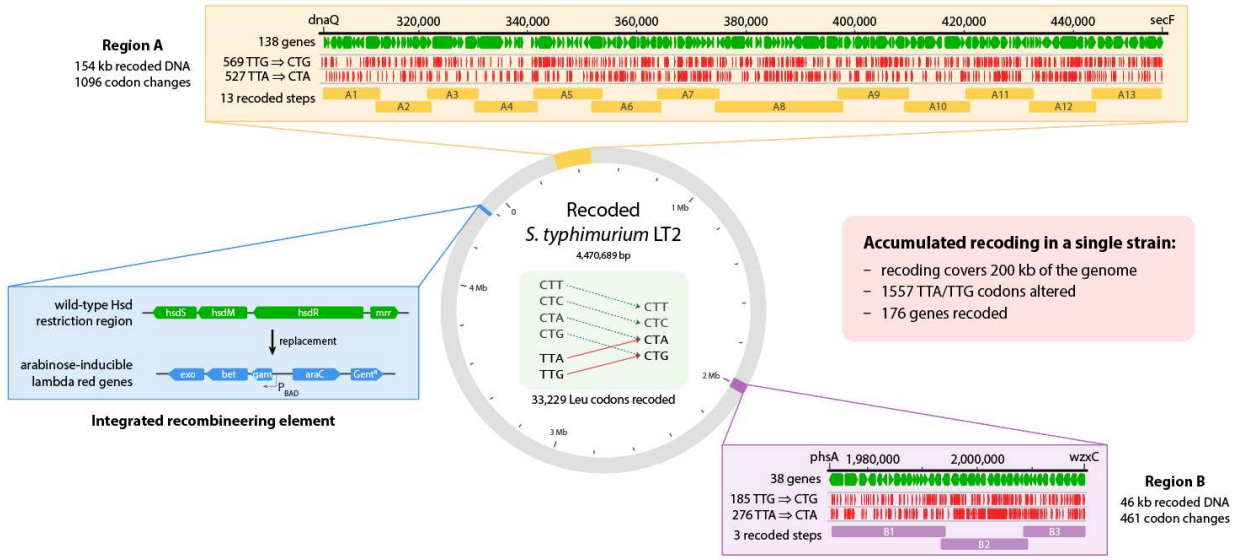

Figure 1. Accumulated recoding covering 200 kb of the *S. typhimurium* genome. An *in silico* design of the recoded genome replaced all 33229 instances of TTA/TTG leucine codons with synonymous CTA/CTG codons respectively. From this design, regions A and B (yellow and purple) were independently recoded using SIRCAS, then combined into one strain by conjugative assembly. The 'integrated recombineering element' (inserted at the native *hsd* restriction system of *S. typhimurium*, blue) is an arabinose-inducible lambda red cassette that facilitates recoding by enhancing the homologous replacement of wild-type genomic DNA with synthetic fragments.

2.3 Results

Computational design of a leucine-recoded *S. typhimurium* genome

A fully recoded *S. typhimurium* genome was generated computationally (Figure 1 and Supplementary Figure SI 1 Section S2), replacing all 33,229 instances of leucine codons TTA and TTG in open reading frames (McClelland et al., 2001) with synonymous CTA and CTG codons respectively. The decision to target leucine codons was due to their high frequency of occurrence throughout the genome (Supplementary File SI 1, Section S2.2), and similar to previous recoding efforts (Ostrov et al., 2016; Wang et al., 2016), the two specific codons were chosen because their corresponding tRNA anticodons are not involved in decoding the remaining four leucine codons (Giege et al., 1998). This recoding

scheme also minimizes the number of base pair changes, requiring only a single T to C base change for each codon. In cases where target codons were present in overlapping reading frames, synonymous changes in both proteins were made where possible. In 49 instances where synonymous mutations were not possible, point mutations were chosen to minimize the impact on biological function as predicted by PROVEAN (Choi et al., 2012), an alignment-based tool for predicting the effect of amino acid substitutions and indels (Supplementary File SI 1 Table S2.3.1). There was one specific instance in which overlapping genes were split due to the presence of four target codons within the overlap (STM0521 and STM0522, see Supplementary File SI 1 Section S2.3). In addition to codon changes, 400 kb of pseudogenes, mobile elements and pathogenicity islands were removed to reduce genetic instability, genome size and pathogenicity (McClelland et al., 2001) (Supplementary File SI 1 Table S2.4.1). Finally, 754 instances of the recognition site for restriction enzyme LguI (5 -GCTCTTC/GAAGAGC-3) were removed to facilitate downstream cloning.

Construction of an *S. typhimurium* recombineering strain

We used a recombineering approach based on the lambda red system for increasing integration efficiency (Datsenko & Wanner, 2000; Sawitzke et al., 2007; Yu et al., 2000). Recombineering is routinely used in *S. typhimurium* and other bacteria for small insertions and knockouts (Karlinsey, 2007), but the efficiency decreases with increasing DNA insert size (Kuhlman & Cox, 2010). Although the use of endonucleases in tandem with recombineering has been shown to improve integration efficiency (7 kb with I-SceI (Kuhlman & Cox, 2010) and 100 kb insertions with Cas9 (Wang et al., 2016), we postulated that these extra accessory proteins are not necessary to achieve practical rates of integration sufficient for genome recoding, even when working with large DNA segments.

To generate a stable recombineering strain of *S. typhimurium*, we generated a construct containing the lambda red genes under arabinose-inducible control, a gentamicin resistance cassette, and homology arms for replacing the *hsd* region of the genome (Figure 1 and Supplementary File SI 1, Section S3.5). Using an *S. typhimurium* strain containing the recombineering plasmid pKD46 (Datsenko & Wanner, 2000; Slechta et al., 2003), the P_{ara}-lambda red cassette was successfully integrated into the *hsd* locus, simultaneously removing the native *hsd* restriction system (Van Pel & Colson, 1974) which could otherwise impede transformation. Curing the pKD46 plasmid at elevated temperature resulted in the final restriction deficient strain containing an integrated recombineering element (IRE).

Preparation of large recoded DNA segments via rolling circle amplification

From the *in silico* genome design, sixteen recoded segments were constructed (A1–A13 and B1–B3, Figure 1 and Supplementary File SI 3) constituting two separate arbitrary regions of the recoded genome (Regions A and B). Each segment contained 10–25 kb of recoded DNA, a selection marker, and 1 kb flanking homology regions for integration. The 10–25 kb size range was chosen to simplify construction, decrease the likelihood of unwanted internal recombination events, and minimize the cost of fixing an error in any one segment. Each segment was assembled in yeast (Gibson et al., 2008) from commercially synthesized 2–4 kb DNA fragments (Figure 2A). The cloned assemblies in yeast were checked by diagnostic digests, but were otherwise used immediately without further sequencing. To assess the utility of different commercial DNA sources, a mixture of 156 kb of clonal sequence-verified DNA and 44 kb of non-clonal DNA was used for constructing the segments (Supplementary File SI 2, Table S1). Clonal DNA refers to sequence-verified constructs propagated through a bacterial host, while non-clonal DNA constructs are synthesized completely *in vitro* in a faster and cheaper process to generate a heterogeneous pool containing errors derived from chemical synthesis. SIRCAS uses a marker

swapping approach alternating between chloramphenicol and kanamycin selection (Figure 2B) for a simple phenotypic readout, similar to the strategy for building chromosomes in *S. cerevisiae* (Annaluru et al., 2014; Richardson et al., 2017) and combining genomes in *B. subtilis* (Itaya et al., 2005).

We bypassed the use of bacterial plasmids, site-specific enzymes (eg. site-specific integrases, Cas9), and associated cloning steps by amplifying each recoded segment directly from yeast using rolling circle amplification (Dean et al., 2001). RCA utilizes the bacteriophage ø29 DNA polymerase and random hexamers to selectively amplify circular DNA. Linearizing the resulting concatemer by LguI digestion gave microgram quantities of DNA ready for direct integration (Figure 2B). Carrying each segment on a

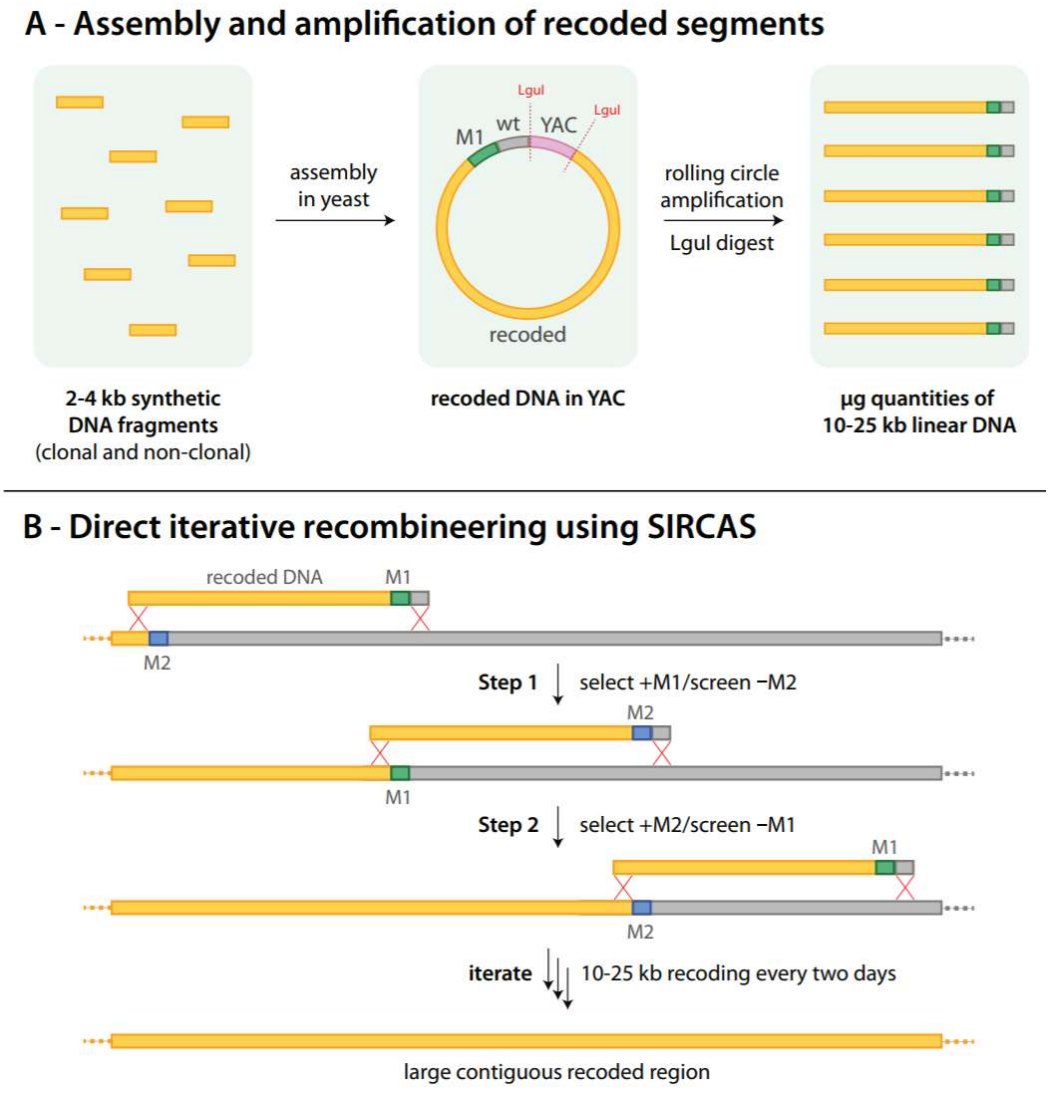

Figure 2. (**A**) Large 10–25 kb recoded DNA segments were created by assembling short synthetic DNA fragments into a YAC, followed by rolling circle amplification and linearization. Each recoded segment contains a selection marker (M1 or M2, typically kanamycin or chloramphenicol resistance cassettes) and flanking homology regions for integration. (**B**) Accumulated genome recoding by SIRCAS involves iterative recombination of recoded DNA segments.

bacterial plasmid would other-wise have required additional negative selection against the plasmid backbone to distinguish between the desired integration event versus the plasmid existing as an extrachromosomal replicative element (Ostrov et al., 2016; Wang et al., 2016). Propagation through a bacterial cloning host may also lead to toxicity issues arising from unintended expression of *Salmonella* genes.

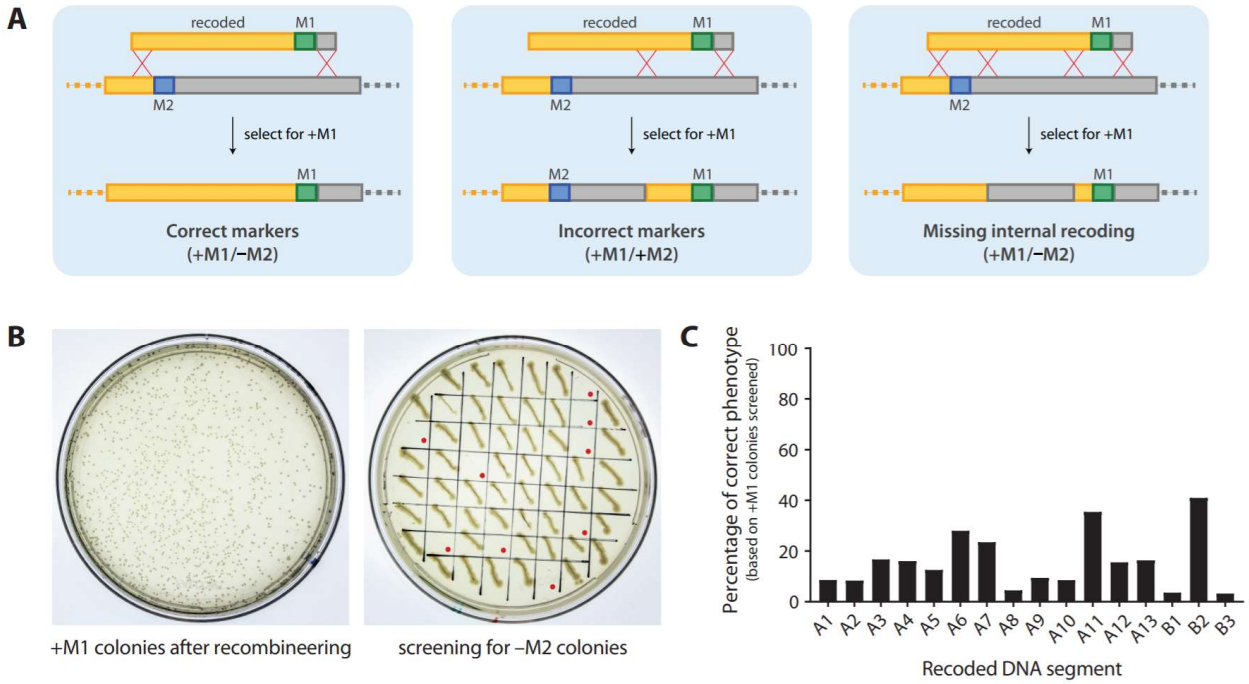

Figure 3. Screening for stepwise replacement of genomic segments. (**A**) Three possibilities after recombineering and selection for marker M1 are full integration of the entire segment leading to marker swapping, partial integration leading to the presence of both markers, or internal crossovers leading to missing recoding. (**B**) Hundreds of transformants were typically obtained on M1 selection plates (recoding of segment A13 shown here). On average, one in six colonies (16%) had the correct phenotype +M1/–M2, as identified by screening for no growth on M1+M2 plates. In this instance, 9 out of 60 colonies were correct (each marked with a red dot). (**C**) The rate of correct marker phenotype varied between different rounds of SIRCAS.

Accumulated genome recoding by SIRCAS

We successfully performed SIRCAS on all sixteen recoded DNA segments to accumulate recoding in two independent *S. typhimurium* strains, one recoded from segment A1 to A13 and the other from B1 to B3. Each round of SIRCAS required only two days to complete. In a typical round of SIRCAS, hundreds of marker-positive colonies were obtained after recombineering (Figure 3). No colonies were obtained in negative controls in which arabinose or DNA was omitted. Upon screening for loss of the previous marker, a median of 14% of colonies was found to have correctly swapped markers, while the remaining colonies incorrectly contained both selection markers (Figure 3B and Supplementary File SI 1, Table S4.2.1). This frequency is higher than naively predicted based on lengths of DNA in the distal and central segments, but is expected based on the results of Matic *et al.* (Maticet al., 1995), who found that mismatches in recombining DNA can significantly suppress the isolation of recombinants. The rate of obtaining the correct marker phenotype ranged from 3–41% (Figure 3C), presumably due to differences in marker integration locus (Englaender et al., 2017), as well as the size and content of the incoming recoded DNA. To check whether recombinants with the correct markers were missing internal recoding due to multiple crossovers (Figure 3A), selected colonies with the correct phenotype were Sanger sequenced at one location in the middle of each newly recoded segment before progressing to the next round of SIRCAS. Correct recoding was observed in 83% of all sequenced colonies with the correct marker phenotype over sixteen rounds of SIRCAS, indicating that multiple internal crossovers do occur but are relatively uncommon (Supplementary File SI 1 Table S4.3.1). This inference was supported by whole genome sequencing of the final organism containing all 200 kb of recoded segments.

Hierarchical assembly of SIRCAS-recoded regions by conjugation

After complete recoding of regions A and B by SIRCAS in two separate strains, the two regions of recoding were consolidated into one strain by conjugative assembly (Isaacs et al., 2011). An origin of

transfer from the Gram-negative bacterial RK2 plasmid, which directs RK2-mediated conjugal transfer of DNA to a broad range of hosts (Guiney & Yakobson, 1983), was integrated upstream of the recoded region of strain A13 by recombineering using an associated spectinomycin resistance cassette (Supplementary File SI 1 Figure S5.2.1). This donor strain was transformed with the 52.7 kb conjugation plasmid pTA-Mob (Strand et al., 2014), which contains the RK2 parABCDE ('partitioning') genes necessary for conjugation. The B3-based recipient strain was prepared by integrating an ampicillin resistance cassette at the start of region A such that upon conjugation, the resistance marker should be lost. Conjugation between these strains resulted in the transfer of the entire 154 kb recoded segment from the donor strain to the recipient, resulting in an exconjugant strain A13-B3 containing a total of 200 kb of recoded genomic DNA. Conjugation success rate as determined by ampicillin marker loss was 40% (17 out of 42 colonies screened).

Characterization of accumulated recoding in *S. typhimurium*

In the final strain (designated A13-B3), recoding throughout both targeted regions spanning 200 kb of the genome was confirmed by next-generation sequencing, with a total of 1557 leucine codon changes successfully installed. An additional five point substitutions, one single point deletion and one single point insertion were found across 156 kb of the recoded regions constructed using clonal sequence verified DNA, constituting an error rate of approximately 1 in 20 000. This number is consistent with the error rate of RCA (Hutchison et al., 2016) (approximately 1 in 50 000). In contrast, a total of 51 errors were found across the 44 kb of recoded regions constructed using non-clonal DNA, for an approximate error rate of 1 in 860 (Supplementary File SI 1 Section S6.3 and Supplementary File SI 2 Tab S7). The majority of these are single base deletions, consistent with expected errors in the synthesis of the chemical oligonucleotide precursors used to generate non-clonal DNA, which occur at a rate of

better than 1 in 200 (Kosuri & Church, 2014). Despite these errors (Supplementary File SI 2 Tab S8), the final strain A13-B3 was still viable and no other instances of incorrect leucine codon recoding were found. Growth rates of all recoded strains were measured to analyse the impact of accumulated recoding on the viability of *S. typhimurium*. Despite the vast number of changes introduced as part of the recoding process, no major cumulative growth defect was observed over the sixteen rounds of SIRCAS, demonstrating that bacterial genomes appear to be amenable towards large-scale rational reengineering. No downward trend in fitness was observed across the entire 200 kb of recoding, with comparable doubling times measured across all strains at 37° C in LB (Figure 4A). Similar trends in growth behaviour were also seen with growth at both 30° C and 42° C (Supplementary File SI 1 Figure S7.1.3). Sequencing also did not reveal any clear compensatory mutations in non-recoded regions of the genome which might suppress any defects (Supplementary File SI 2 Table S9).

Minor phenotypic changes were observed at one step of the recoding process. When grown in a plate reader, cells containing segment B2 (ie. strains B2, B3 and A13-B3) were observed to sediment unevenly (Figure 4B), resulting in anomalous growth curves. Growth experiments repeated in Erlenmeyer flasks showed no major differences in doubling time (Figure 4A). Although final optical density measurements differed between strains, growth as measured by colony forming units was consistent between wild-type and recoded strains B2, B3 and A13-B3 (Figure 4C). The sedimentation effect may be due to changes in the cell surface arising from non-clonal DNA errors in regions responsible for *O*-antigen biosynthesis (Supplementary File SI 1 Figure S6.3.1). No stepwise accumulation of growth defects was observed upon recoding all other segments.

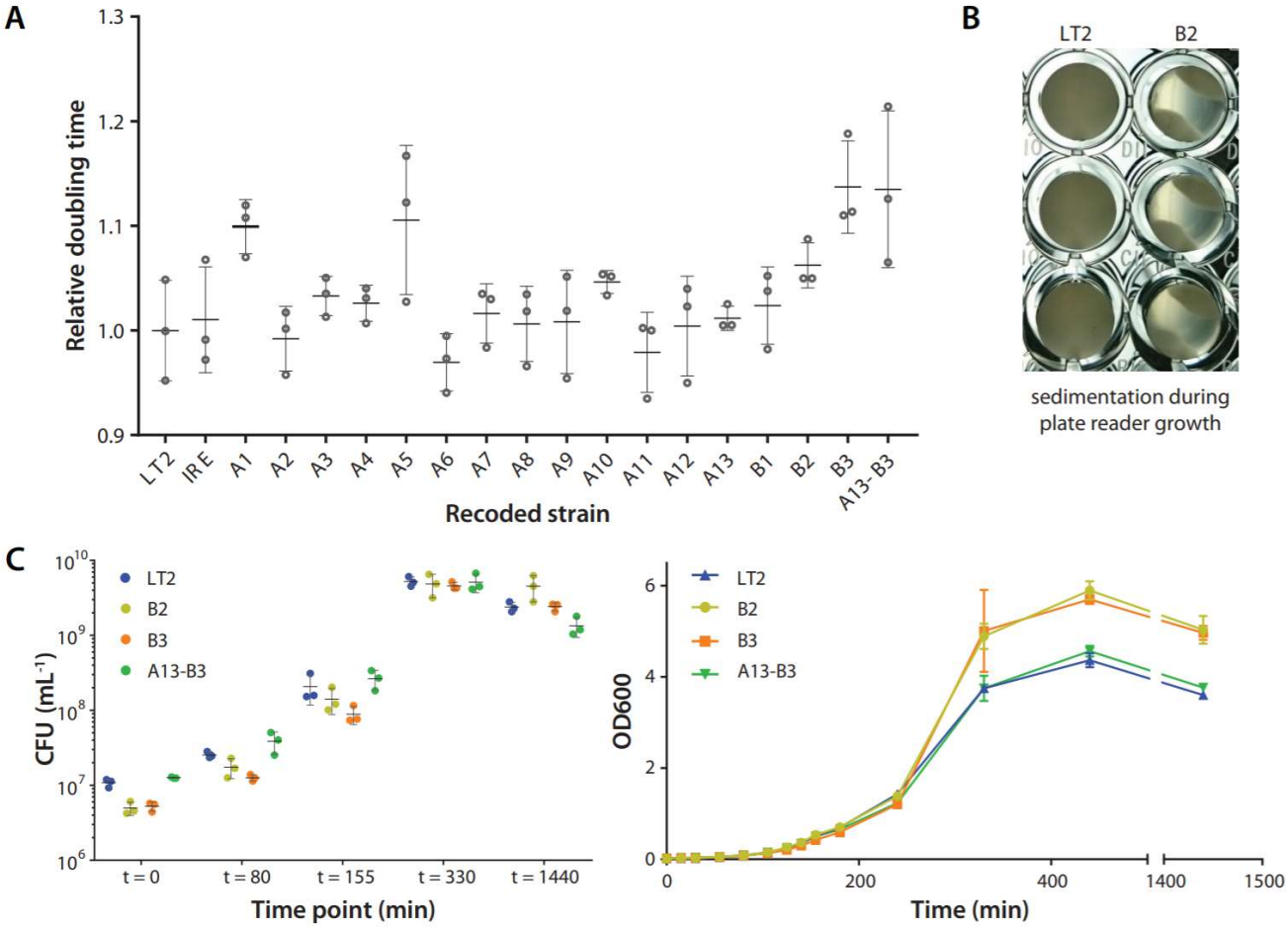

Figure 4. (**A**) Doubling times of various recoded *S. typhimurium* strains growing at 37 ºC in LB. Each data point is the average of three technical replicates, and the error bars represent the standard deviation of three biological replicates. No major fitness defect was observed throughout the recoding process. The strain nomenclature 'A ' refers to the strain containing all recoded segments from A1 up to A . We note that doubling times for B2, B3 and A13-B3 were calculated from growth curves conducted in batch cultures grown in flasks due to sedimentation, while the remaining data was obtained on a plate reader. (**B**) Uneven sedimentation of cells is observed when strains B2, B3 and A13-B3 are grown in a plate reader (wells in right column). (**C**) Growth data for B2, B3 and A13-B3 was obtained by growth in culture flasks rather than in 96-well plates to avoid sedimentation artifacts. Wild-type LT2 was used as a control and had comparable doubling times in both plate and flask growth (see Supplementary File SI 1 Figure S7.1.4).

2.4 Discussion

We have developed SIRCAS (stepwise integration of rolling circle amplified segments) as a rapid method for making genomic changes that bypasses the requirement for site-specific enzymes. Applying

SIRCAS to recoding of the *S. typhimurium* genome, we installed 1557 TTA/TTG to CTA/CTG leucine codon changes across 200 kb of genomic DNA in a single strain. A key aspect of SIRCAS is the use of rolling circle amplification to produce DNA of sufficient quality and quantity appropriate for transformation. We found that recombineering with extended homology regions was effective on 10–25 kb linear DNA segments, without needing assistance from extra accessory proteins such as endonucleases. Sixteen cycles of integration into the genome were achieved efficiently, with each integration step requiring two days to complete. The recoded strains did not show any major fitness defects when compared to the wild-type strain, demonstrating the remarkable plasticity of bacterial genomes. Overall, this work paves the way for creating a completely recoded *Salmonella* as a genetically isolated chassis for downstream therapeutic applications.

The SIRCAS approach can be carried out in parallel in many strains for more rapid construction. In this study, we maintained two different recoded strains and unified the recoded regions by conjugation. For completing the construction of a fully recoded *S. typhimurium* genome, we anticipate recoding 8–16 strains in parallel, followed by hierarchical assembly via conjugation. Recoding 20 kb every two days, it is theoretically possible to rewrite 300 kb of the genome per strain in a month.

The *Salmonella* genome tolerated a vast number of designed codon changes, suggesting that recoding the entire genome of *Salmonella* will be feasible. This result is consistent with similar recoding studies in *E. coli* (Ostrov et al., 2016; Wang et al., 2016). Design flaws may affect viability, and troubleshooting may be a significant challenge when constructing a complete recoded genome. As observed in the 'minimal genome' work by Venter *et al.* (Hutchison et al., 2016), synthetic lethal gene pairs may become apparent as more recoding is accumulated in a single strain. In these cases, the piecewise nature of the SIRCAS strategy may prevent the process of redesigning problematic pieces

from becoming a bottleneck. Further understanding of principles governing codon choice will also help improve our existing designs (Kelsic et al., 2016; Napolitano et al., 2016).

There are currently few alternative methods for large-scale genomic engineering. Multiplex automated genome engineering (Wang et al., 2009) and multiplexed CRISPR/Cas9 systems (Cong et al., 2013) are not sufficiently high throughput to install the large number of designed changes required for leucine recoding. Church and co-workers used site specific integrases for testing a 57-codon *E. coli* genome design (Ostrov et al., 2016), integrating 10 out of 87 possible 50 kb recoded segments into independent strains. Site-specific integrases are an effective means for testing individual recoded segments, as there is less opportunity for internal crossover events when compared to homologous recombination. However, unification of recoded segments into a single strain is more challenging. Integrase based methods require the installation and subsequent removal of landing pads for each integration step. The integration event itself also requires several steps, starting with introduction of the recoded segment on a plasmid, removal of the wild-type genomic segment, integration of the recoded segment, and finally removal of the plasmid. This multistep approach compromises the efficiency of the construction process when compared to the single-step integration of SIRCAS.

The 20 kb size range may be an optimal compromise between the number of integration steps and ease of construction and troubleshooting. Chin and co-workers have combined Cas9 and recombineering to insert over 100 kb of DNA into the *E. coli* genome (Wang et al., 2016). The major advantage of using large 100 kb constructs is that fewer rounds of integration are required to achieve complete genome recoding when compared to the 10–25 kb segments we used for SIRCAS. Conversely, there is a greater likelihood of incomplete recoding due to multiple internal crossover events with increasing size. Larger constructs are also more difficult to assemble from small fragments (efficiency drops as more fragments are included in the assembly) (Gibson et al., 2008) and more prone to shearing

when handled. Furthermore, finding the causative mutations when recoding changes impact fitness becomes more challenging with increasing size. A noteworthy troubleshooting method was described by Chin and co-workers, using a sequencing approach to pinpoint where single lethal mutations arise. However, this method was only demonstrated for recoding a 20 kb region of the *E. coli* genome, and may be more problematic in segments with more than one lethal mutation.

The overall error rates of SIRCAS are competitive with that of other genome recoding methods. Within the 156 kb of recoded regions that was written with clonal sequence-verified DNA, an overall error rate of 1 in 20 000 was observed (7 point mutations and one leucine codon reversion in 200 kb). In comparison, Church and co-workers (Ostrov et al., 2016) reported an error rate of 1 in 5000 (an average of 9.7 mutations and 0.6 codon reversions in 50 kb). For the 44 kb of recoded regions written using non-clonal DNA, an overall error rate of 1 in 860 was found (51 errors, primarily single point deletions). Use of clonal DNA for SIRCAS was therefore preferable, although issues associated with construct errors in non-clonal DNA may become less significant in the future as DNA synthesis and error-correction methods continue to improve.

Beyond genome recoding, SIRCAS is a powerful enabling technology for building and testing large *de novo* designs that span hundreds of genes. SIRCAS is not limited to *Salmonella*, and could be used in any recombineering competent host to integrate large naturally occurring or *de novo* designed gene clusters, facilitating the development of new industrial production strains. SIRCAS also enables interrogation of bacterial genetics on a scale that is not possible with traditional techniques. By making genome scale construction more accessible, SIRCAS will help to empirically elucidate the underlying principles for completely *de novo* genome design.

Acknowledgements

We thank Eric Kofoid and Natalie Duleba from the laboratory of Prof. John Roth (Department of Microbiology, University of California Davis) for the kind gift of the recombineering *Salmonella* strain TT22971 containing plasmid pKD46. We also thank Prof. George Church (Department of Genetics, Harvard Medical School) and members of his laboratory for general discussions about genome recoding.

Author contributions: The overall study was conceived by P.A.S. and J.C.W. Experimental design was conducted by Y.H.L., F.S., J.K., M.A.P.K. and A.R. The computational design of the recoded genome and other computational analyses were performed by M.A.P.K., A.R. and Y.H.L. Assembly and RCA of DNA constructs was performed by Y.H.L., F.S., C.A.H., E.S. and D.L. The protocols for DNA assembly and RCA were established through guidance and troubleshooting from M.T.W., D.G.G. and Y.A.C. Genomic integration (SIRCAS) was conducted by Y.H.L. and C.A.H. Conjugation experiments were performed by J.K. Sequence analysis was conducted by Y.H.L. and F.S. Growth data for recoded strains was obtained by Y.H.L. All authors contributed to writing the manuscript.

Funding

DARPA [BRICS HR0011-15-C-0094]; Wellcome Trust for the award of a Sir Henry Wellcome postdoctoral fellowship [107402/Z/15/Z to Y.H.L.]; acknowledges funding from a DOE CSGF fellowship [DE-FG02-97ER25308 A.R.]; NIH training grant [5T32GM007598 F.S.]. Funding for open access charge: BRICS [HR0011-15-C-0094].

Conflict of interest statement. D.G.G. is co-founder of SGI-DNA and Vice President of DNA Technologies at Synthetic Genomics, Inc., and M.T.W. is an employee of Synthetic Genomics, Inc.

2.5 Materials and Methods

Assembly of synthetic DNA in yeast

Synthetic DNA fragments were either purchased from commercial vendors (Gen9, SGI-DNA and Integrated DNA Technologies), or obtained by automated enzymatic assembly of oligonucleotides (BioXp™ 3200 instrument, SGI-DNA). Full details of construct design are provided in the Supporting Information (Supplementary File SI 1 Section S3).

The DNA assembly method by transformation-associated recombination was adapted from previously reported protocols (Gibson et al., 2008; Kouprina & Larionov, 2008). To prepare yeast spheroplasts for transformation, *Saccharomyces cerevisiae* VL6-48 (ATCC® MYA3666™) was grown in 50 ml YPAD medium to an optical density OD_{600} of 1. Cells were then suspended in 20 ml of 1 M sorbitol and kept at 4° C for 4 h. Spheroplasts were generated in 20 ml SPE buffer containing 20 ml beta-mercaptoethanol (Sigma-Aldrich) and 250 mg Zymolyase-20T (US Biological) at 30° C. When the OD_{600} of cells diluted 1:10 in SPE buffer was 3–4 times the OD_{600} value of cells diluted in 2% SDS in SPE buffer, the cells were washed twice with 50 ml 1 M sorbitol. Cells were then resuspended in 2.5 ml STC buffer. After 20 min, 200 ml of spheroplasts were added to a 50 ml solution of DNA fragments (total 1 mg DNA, equimolar fragments). After 10 min, 1 ml of freshly prepared 20% PEG 8000 solution was added. After a further 20 min, the spheroplasts were resuspended in 800 ml SOS and incubated at 30° C for 30 min. The recovered cells were then suspended in top agar kept at 55° C and poured onto synthetically defined minus Trp plates. Buffer and media formulations are detailed in the Supporting Information (Supplementary File SI 1 Section S1). Successful assemblies were identified either by diagnostic digests of RCA products or multiplex PCR across assembly junctions (Supplementary File SI 1 Section S3.6).

Rolling circle amplification of DNA from yeast artificial chromosomes

Adapting the protocol by Hutchison *et al*. (Hutchison et al., 2016), yeast centromeric plasmid DNA was isolated from *Saccharomyces cerevisiae* using a modified miniprep workflow. Overnight 5 mL cultures of yeast were suspended in 250 ml P1 buffer (Qiagen) containing 250 mg Zymolyase-20T (US Biological). After incubation at 37° C for 1 h, 250 ml P2 buffer and 250 ml P3 buffer (Qiagen) were added sequentially. The plasmid DNA in the supernatant was precipitated with 1 ml isopropanol, washed with 1 ml of 70% ethanol, and the resulting DNA pellet was airdried. The plasmid DNA was then dissolved in 40 ml Tris–EDTA buffer pH 8.

Rolling circle amplification was conducted using reagents from the TempliPhi Large Construct Kit (GE Healthcare). 2 ml of plasmid DNA solution was added to 15 ml of sample buffer and heated at 95° C for 3 min. Upon cooling, 15 ml of reaction buffer, 1 ml enzyme and 2 ml of 10 mM dNTPs (New England Biolabs) were added, and the reaction was incubated for 24 h at 30° C.

To release the recoded DNA as a linear construct, the amplified DNA was diluted with 40 ml water, then 8 ml of 10× FastDigest buffer and 2 ml LguI restriction enzyme (Thermo Fisher) was added. After 1 h at 37° C, the linearized DNA was precipitated with ethanol and redissolved in 40 ml Tris–EDTA buffer pH 8.

Integration into the *Salmonella* genome

Overnight cultures of *S. typhimurium* were diluted 1:100 into a 50 ml volume in LB containing 10 mM L-arabinose (Sigma-Aldrich). When an OD_{600} of 0.5 was reached, cells were washed twice with 20 ml and once with 1 ml of ice-cold 10% glycerol, then resuspended in 250 ml of 10% glycerol. 50 ml of the competent cell suspension was electroporated with the appropriate DNA construct (100 ng per kb of construct, Bio-Rad MicroPulser using the Ec2 setting) in a 0.2 cm cuvette (USA Scientific). After 5 h outgrowth at 37° C in 1 ml SOC medium, 250 ml of cells was plated onto antibiotic selection plates.

Typically, at least 50 colonies were patched onto double antibiotic selection plates (chloramphenicol and kanamycin) to determine which colonies contained the correct markers. See the Supporting Information for further details about the SIRCAS protocol (Supplementary File SI 1, Section S4).

Conjugative assembly of recoded genomic regions

Donor strains were prepared by integrating a cassette containing spectinomycin resistance and the origin of transfer from RK2 plasmid (Guiney & Yakobson, 1983) upstream of recoded region A by recombineering, and transforming plasmid pTA-Mob (Strand et al., 2014) containing necessary parABCDE ('partitioning') genes. Recipient strains were prepared by integrating an ampicillin resistance cassette at the start of the recoding region to be transferred from the donor.

Conjugation experiments were based on a protocol previously described by Ma, Moonan and Isaacs (Isaacs et al., 2011). Donor and recipient strains were grown in separate 30 mL cultures, inoculated from overnight cultures by 1:100 dilution. The recipient strain lambda red system was induced with arabinose to enhance homologous recombination. Cultures were grown to OD_{600} 0.5, centrifuged and re-suspended in LB to OD_{600} 15. Donor and recipient cells were mixed in 80– 120 ml aliquots and pipetted as 10–20 ml spots on LB plates, incubated at 30° C for 1–2 h, then washed and plated as 250 ul of a 1:100 dilution on selection plates. Candidates were patched out on multiple selection plates to check for correct phenotype and incubated at 37° C.

Next-generation sequencing of recoded strains

Salmonella genomic DNA was prepared using a DNeasy Blood and Tissue kit (Qiagen) and quantified using a Qubit 2.0 fluorometer (Thermo Fisher). DNA libraries for next-generation sequencing were prepared using a Nextera XT kit (Illumina) with size selection using SPRIselect beads (Beckman-

Coulter) according to manufacturer's protocols. Sequencing was conducted on a Miseq platform using either a MiSeq or MiSeq Nano reagent kit v2 (Illumina), running 300 cycles with paired ends. Reads were trimmed and aligned to a reference genome using the Geneious software package (version 9.1.5) (Kearse et al., 2012). Further details can be found in Supplementary File SI 1 Section S6.

Measurement of recoded *Salmonella* growth rates

Cells were inoculated into 200 ml LB in a 96-well clear flat-bottom plate (Corning) and incubated overnight at 37° C with continuous shaking. From these overnight stationary cultures, 1 ml was added to 200 ml fresh LB in a new 96-well plate, and incubated at 37° C with continuous shaking in a plate reader (BioTek Synergy HT). OD_{630} measurements were taken every 10 min for at least 16 h. Measurements were performed in technical triplicates, and at least three biological replicates were obtained for each sample. Errors are calculated as the standard error of the mean between the biological replicates.

Doubling times were determined in a similar manner to Lajoie *et al*. (Lajoie et al., 2013), based on linear regression of ln (OD_{630}) using five adjacent time points (40 min). The doubling time was calculated by $t_d = \ln(2)/m$, where m is the maximum gradient of $\ln(OD_{630})$ as determined by the linear regression analysis. Additional growth data can be found in Supplementary File SI 1, Section S7.1. For strains B2, B3 and A13-B3, cells were inoculated into 5 ml LB cultures overnight. From these overnight stationary cultures, 250 ml was added to 25 ml fresh LB in a 250 ml unbaffled Erlenmeyer flask, and grown in a Multitron Standard (INFORS HT) shaking incubator at 37° C and 200 rpm. OD_{600} measurements were taken every 15–25 min during early exponential phase on an Ultrospec 10 (Amersham Biosciences), removing an aliquot of 700 ml each time. Doubling times were calculated using the same procedure except using three adjacent timepoints which had OD values lying between 0.05 and 0.75. Colony forming unit (CFU) measurements were performed by counting colonies from

appropriate dilutions on LB plates. As a control, growth curves for wild-type LT2 were obtained in both plate reader and Erlenmeyer flask format, and the doubling times were found to be consistent with each other.

Chapter 3 – Rational design of evolutionarily stable microbial kill switches

Finn Stirling[1,2], Lisa Bitzan[1], Samuel O'Keefe[1], Elizabeth Redfield[1], John WK Oliver[1,2], Jeffrey Way[1,2], and Pamela A. Silver[1,2,3]

[1] Department of Systems Biology, Harvard Medical School, 200 Longwood Avenue, Warren Alpert 536, Boston, MA 02115, USA

[2] Wyss Institute for Biologically Inspired Engineering, Harvard University, 3 Blackfan Circle, 5th Floor, Boston, MA 02115, USA

[3] Lead Contact

Correspondence: pamela_silver@hms.harvard.ed

This article was first published in Molecular Cell, vol. 68, 686-697, copyright Elsevier (2017)
doi: 10.1016/j.molcel.2017.10.033

My contribution

For the essentializer element, the initial project design was carried out before I joined the Silver lab. The experiments behind the screening process were conceived of by Jeffrey Way, but entirely implemented by me. All other aspects of the project, including overall plan, construct design and ordering, construction of strains and all experimental analysis was carried out by me or under my direct supervision by the three students working with me; Lisa Bitzan, Samuel O'Keefe and Elizabeth Redfield. The final manuscript was written almost exclusively by me with edits from the other authors, and additions from Jeffrey Way for the discussion of evolutionary stability.

3.1 Abstract

The evolutionary stability of synthetic genetic circuits is key to both the understanding and application of genetic control elements. One useful but challenging situation is a switch between life and death depending on environment. Here are presented "essentializer" and "cryodeath" circuits, which act as kill switches in *Escherichia coli*. The essentializer element induces cell death upon the loss of a bi-stable cI/Cro memory switch. Cryodeath makes use of a cold-inducible promoter to express a toxin. We employ rational design and a toxin/antitoxin titering approach to produce and screen a small library of potential constructs, in order to select for constructs that are evolutionarily stable. Both kill switches were shown to maintain functionality in vitro for at least 140 generations. Additionally, cryodeath was shown to control the growth environment of a population, with an escape frequency of less than 1 in 10^5 after ten days of growth in the mammalian gut.

Keywords:

Containment, Library, Synthetic Biology, Lambda, CspA, Promoter, Toxin, Antitoxin, Cold Shock.

3.2 Introduction

As synthetic biology makes advances in producing real world applications using genetically engineered micro-organisms, the issue of biological containment becomes increasingly important. Safeguards have previously been developed that require the addition of a survival factor to maintain viability in a bacterial population. Approaches include inducing an auxotrophy for a particular metabolite (Gallagher et al., 2015, Steidler et al., 2003), repressing the expression of an essential gene (Cai et al., 2015; Chan et al., 2015, Gallagher et al., 2015) or re-writing the genetic code to ensure dependency on a synthetic amino acid (Rovner et al., 2015). However, future applications of synthetic biology look beyond the confines of a laboratory. Strains have already been developed that can degrade inorganic polymers to reduce waste (Yoshida et al., 2016), provide sustenance or energy during space travel (Menezes et al., 2014, Montague et al., 2012, Way et al., 2011), or that colonize the mammalian gut to help diagnose and treat pathogenic infections (Kotula et al., 2014, Steidler, 2003). For these applications, a new form of containment is required for uncontrolled environments, one that does not require human monitoring or input.

Evolutionary instability is an inherent flaw in any biological control using a lethal effect to control environmental growth. Microorganisms rapidly evolve to remove any genetic element that reduces fitness (Knudsen & Karlstrom, 1991, Molin et al., 1993). “Kill switches” are defined as artificial systems that result in cell death under certain conditions. Several kill switches have been explored for containment of engineered microbes, but necessarily involve lethal genes that are induced in designated non-permissible conditions. Thus any kill switch with leaky, low level expression of a toxin in permissible conditions may be quickly disabled in rapidly growing microbes.

Although a host of effective kill switches have been described, most evolve to lose functionality within days (Chan et al., 2015), or have no data supporting their longevity (Ahrenholtz et al., 1994, Caliando & Voigt, 2015, Callura et al., 2010, Djordjevic et al., 1997, Kong et al., 2008, Piraner et al., 2016). One exception is a multi-layered kill switch (Gallagher et al., 2015), which is stable for at least 110 generations but requires external supplementation of survival factors. Mutational loss of a microbial kill switch occurs in the context of an asexually reproducing population subject to "periodic selection"(Atwood, 1951, Maddamsetti et al., 2015 Novick & Szilard, 1950). In this regime, selection occurs at one locus at a time, and loss of a synthetic-biological device implies that it is the most deleterious element in a genome. To be evolutionarily stable, an engineered element can have a small fitness cost, provided that its selection coefficient is less than 1-10%, the typical fitness advantage associated with mutations that drive adaptive sweeps (Maddamsetti et al., 2015).

To create evolutionarily stable kill switches, we used an approach for varying the level of expression in a toxin/antitoxin system. Small rationally designed libraries were created with key bases modified in the promoter and RBS sites of both toxin and antitoxin. This system is broadly applicable in kill switch designs with diverse forms of regulation. We demonstrate this approach with two unrelated control systems: (1) the early termination of transgenics that lose function of another engineered module; and (2), the temperature-dependent termination of a transgenic microorganism that resides in the mammalian gut. In doing so, we explore a limited rationally designed space to achieve the optimised behaviour of genetic circuits.

3.3 Results

General design strategy

An environmentally sensitive kill switch needs to be highly lethal in non-permissible conditions, but stable enough in permissible conditions to avoid conferring an evolutionary disadvantage. Achieving intended levels of protein repression can be more straightforward when a molecular 'sponge' is included to titrate low levels of an expressed toxin (Bläsi & Young, 1996; Nathan et al., 2016). In our system, this has been achieved through toxin/antitoxin pairings where the antitoxin sequesters toxin expressed during permissible states (Figure 1A), but does not prevent cell death after a non-permissible event. The toxin is placed under the control of an environmentally sensitive promoter, capable of strong repression in permissible conditions and high levels of expression in non-permissible conditions (Figure 1B). Even with highly stringent repression of a promoter, there will always be leaky expression of the intended gene product. For a toxin lethal enough to make an effective kill switch, any unintended expression would have a negative effect on cell survival and decrease evolutionary stability. For this study we utilized the type II toxin-antitoxin system CcdB/CcdA. CcdB is a lethal toxin for many Enterobacteriaceae (Wright et al., 2013). It targets the GyrA subunit of DNA gyrase, arresting it in the intermediate stage of action after a double-stranded break has been induced, resulting in cell death. (Madl et al., 2006).

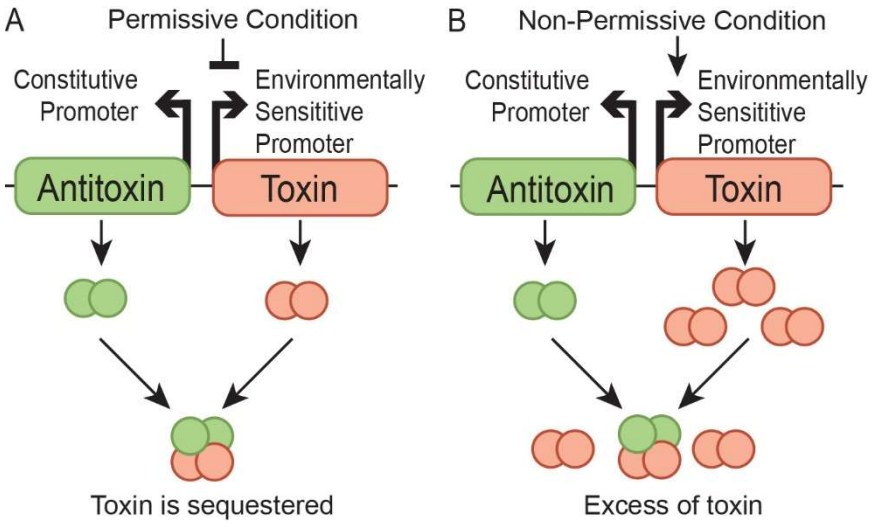

Figure 1: Overview of kill switch design concept A) In permissible conditions the toxin is repressed. The antitoxin is expressed at a constitutive low level to accommodate for any leaky expression of the toxin. **B)** Upon a change of environment to non-permissible conditions the repression is lifted and toxin expression increases. The low level of antitoxin expression is no longer capable of preventing a lethal level of free toxin.

We calculated the probability that expressed [toxin] > [antitoxin] for when the average number of transcription events for each gene varies from 1 to 30 (Figure S6A). To delineate between the mutating and non-mutating regions of the landscape (Figure S6B), we assume that in an asexually reproducing population, only the most strongly deleterious mutation will actually be selected against (Atwood, 1951; Koch, 1974). This results in conservation of sequences that may be mildly deleterious. Such mutations causing adaptive sweeps have selection coefficient of about 2-5% per generation (Koch, 1974; Maddamsetti et al., 2015; Novick & Szilard, 1950). We therefore defined the 'non-selected' region of the kill-switch landscape as when the element incorrectly enters the killing state less than 1% of the time.

An 'Essentializer' Element to Select for the Preservation of a Transgenic Cassette:

In previous work, we have published a bistable genetic switch, the 'memory element', which is able to stably record and maintain an output from an exogenous signal (Kotula et al., 2014). The memory element is required to maintain its function for extensive periods of time, over which synthetic circuits are prone to mutation and deletion (Sleight et al., 2010). The memory element relies upon the bacteriophage lambda transcription factors cI and Cro. cI binds preferentially to the operator regions O_R1 and O_R2, whilst Cro binds preferentially to O_R3. Binding of cI to O_R1 and O_R2 represses the expression of *cro*, and vice versa binding of Cro to O_R3 represses the expression of *cI*.

The essentializer element kill switch was designed using the toxin-antitoxin system CcdB/CcdA, and would select for the presence of the memory element. For the essentializer element, the lambda phage operator sites were reordered and positioned over the lambda P_R promoter such that binding of

either cI or Cro would repress the expression of the toxin (Figure 2A and 2B), but in the absence of expression of both transcription factors the toxin would be expressed (Figure 2C). Furthermore, the binding sites O_R1 and O_R2 were modified to increase the binding affinity of cI (Sarai & Takeda, 1989, Takeda et al., 1989). A modified O_R3 was placed downstream of the -10 region for the P_R/toxin promoter in a position analogous to *lacO* within *lacP*; a protein binding to the major groove of these bases is expected to sterically prevent RNA polymerase binding (Murakami et al., 2002). The operators are separated by 6-7 bases, which should allow cooperative binding of cI to either O_R1-O_R2 or O_R1-O_R3 in this configuration (Ptashne et al., 1980).

A specific design goal was to quantitatively adjust the levels of toxin and antitoxin expression so that when cI and Cro proteins are absent, toxin expression is sufficient to kill the cell, but when either Cro or cI protein is present, the toxin is sufficiently repressed such that cell growth is not affected. This should be true even allowing for stochastic binding of repressor or Cro to the operators and expression of the antitoxin. To achieve this goal, a small rationally designed library of essentializer element candidates was constructed to introduce degeneracy at key locations in the regulatory region of the kill switch (Figure 2D). Three bases were varied in the RBS for the antitoxin, each with 4 different possible nucleotides; two bases were varied in the -10 region of the antitoxin promoter, each with 2 different possible nucleotides; one base was varied in the -35 region of the toxin promoter, with 2 different possible nucleotides; two bases were varied in the RBS for the toxin, one with 2 possible nucleotides and the other with 3. This resulted in a library size of 4^3 x 3^1 x 2^4 = 3072 potential combinations. The promoter variations were designed based on (Mulligan et al., 1984), who estimated that, for example, the possible non-preferred bases in the 4^{th} and 5^{th} positions in the consensus -10 region (TAT**AA**T) cause a mild and roughly equal decrement in promoter strength, so that incorporating only two variants would generate maximal functional variation while limiting the number of candidates to be manually screened. Variations in the toxin RBS were chosen

to potentially preserve cI and Cro-binding at the overlapping O_R3 element (Sarai & Takeda, 1989; Takeda et al., 1989). A construct was also made with a frame shift mutation in the toxin open reading frame to act as a non-lethal control ("EE toxin mutant").

The essentializer element candidate pool yielded constructs that depended upon the memory element for survival. A series of tests was designed to select for candidates that survived in the permissible conditions (presence of memory element in the *cI* state) but failed to grow in non-permissible conditions (absence of memory element) (Figure 2E). In test 1, the essentializer element library was recombined into the lambda *red* expressing strain TB10. TB10 expresses cI^{857}, a temperature sensitive mutant of *cI* (Caulcott & Rhodes, 1986) which is active at 30 ^{o}C but inactive at 42 ^{o}C, therefore repressing *ccdB* in a temperature dependent manner. Any recombinant that failed to grow at 42 ^{o}C but survived at 30 ^{o}C was considered potentially lethal. Test 2 used a P1 transduction of a wild-type MG1655 lysate to remove cI^{857}, allowing the strain to regain an operational biotin operon whilst removing the lambda *red* machinery. Successful growth on minimal media without biotin therefore implied a non-lethal essentializer element as cI^{857} would no longer be present to repress. Test 3 involved P1 transducing each essentializer element candidate into MG1655, relying on the distance between the loci of cI^{857} and the essentializer element to prevent them being co-transduced. With no source of cI or Cro in MG1655, if a candidate failed to produce any transductants, it was a candidate. Test 4 transduced the essentializer elements into PAS132, a strain containing the memory element, as well as an ATc responsive trigger element (Jonathan W Kotula et al., 2014) preset to the cI state. If a candidate produced transductants here, but had failed to produce transductants in test 3, it was a candidate. Candidates were sequenced across the antitoxin,

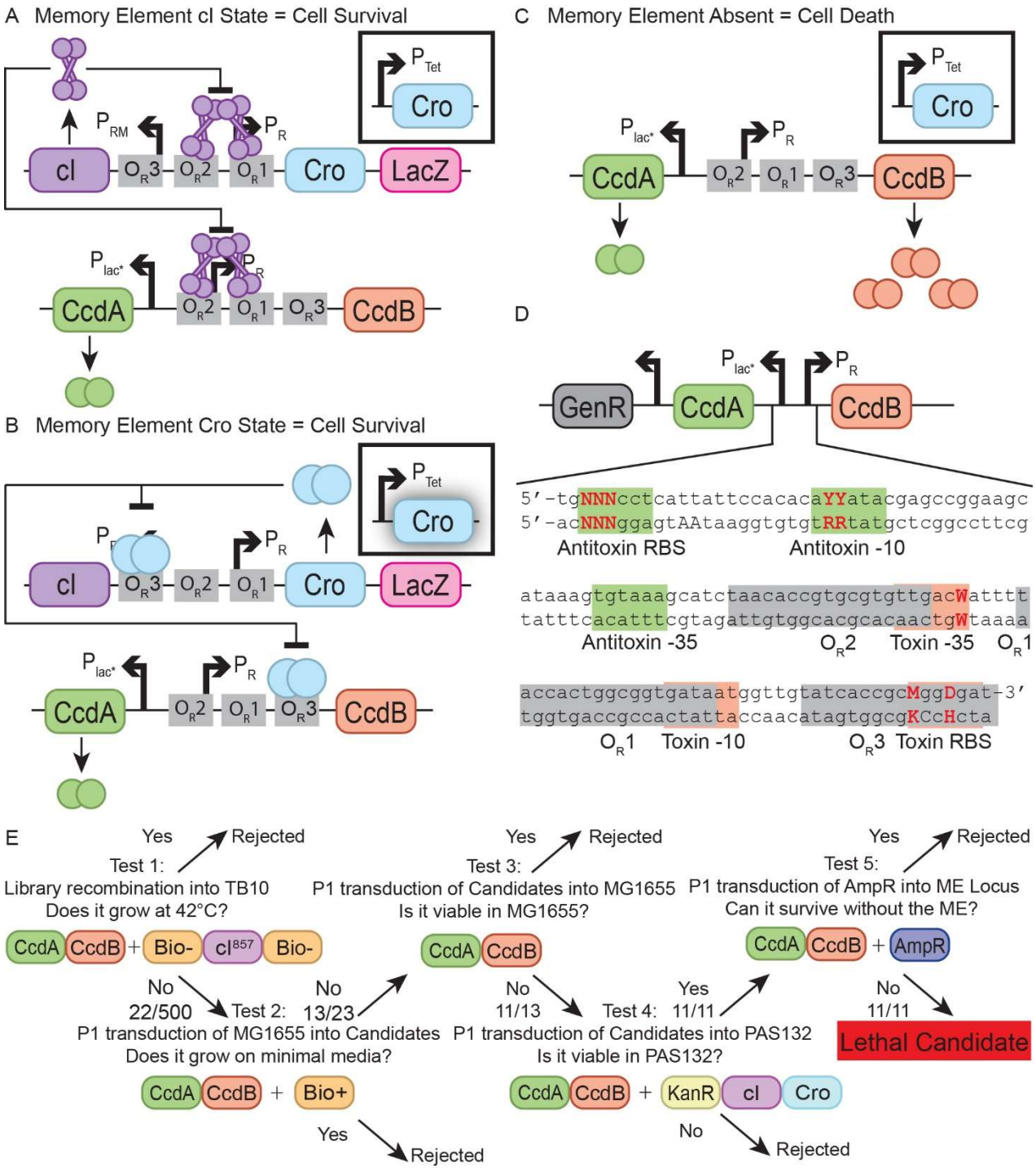

Figure 2: Design and construction of the essentializer element kill switch. A) Expression of *cI* represses expression of *cro* and *lacZ* in the memory element (top cassette), whilst simultaneously repressing expression of *ccdB* in the essentializer element (bottom cassette). *ccdA* is expressed at a constitutive low level. **B)** Exposure to tetracycline leads to a pulse of expression of *cro* from the trigger element (boxed off). Expression of *cro* allows for the expression of *lacZ*, whilst simultaneously

repressing *cI* and *ccdB*. **C)** Memory element is absent. Without repression from *cI* or *cro*, *ccdB* is expressed at lethal levels. **D)** The engineered region encompassing the regulatory DNA for *ccdB* and *ccdA*. Highlighted are loci that have a strong impact on expression level for both the toxin (red) and antitoxin (green), as well as operator binding sites for cI and Cro (grey). Key bases that have been varied are emphasized. N = A, C, T or G; Y = C or T, W = A or T; M = A or C; D = A, T or G. **E)** Overview of the screening process for identifying lethal essentializer element candidates. Displayed are the relevant genotype after each step has been completed and the fraction of candidates that passed a given screen. See also Figure S1.

toxin and regulatory region to determine the exact sequence (Figure S1 and Tables S1-S2), and termed candidates EE01-EE11. All other sequenced candidates were termed EE12- EE31.

The new candidate strains containing the essentializer element in a PAS132 background showed selection for maintaining the memory element. The fifth test inserted ampicillin resistance by P1 transduction into the candidate strains approximately 3000 bp distal to the memory element, likely replacing it. The more common transductant result of removing the memory element would result in only ampicillin resistance and the essentializer element remaining, but a rare transduction event could insert ampicillin resistance without removing the memory element and its associated kanamycin resistance cassette (Figure 3A). If a candidate contained a functional essentializer element, it would not be able to survive without the memory element. In this case only the rare transductant would be viable, whereas controls would have a majority of common transductants with a low proportion of rare transductants. In all 11 candidates, only the rare transductant result was recovered out of 52 transductants screened for growth on ampicillin and subsequently restreaked onto kanamycin (Figure 3B red bars). It should be noted that this assay only tests for the expression of either cI or Cro, and does not account for other potential failures of the memory element.

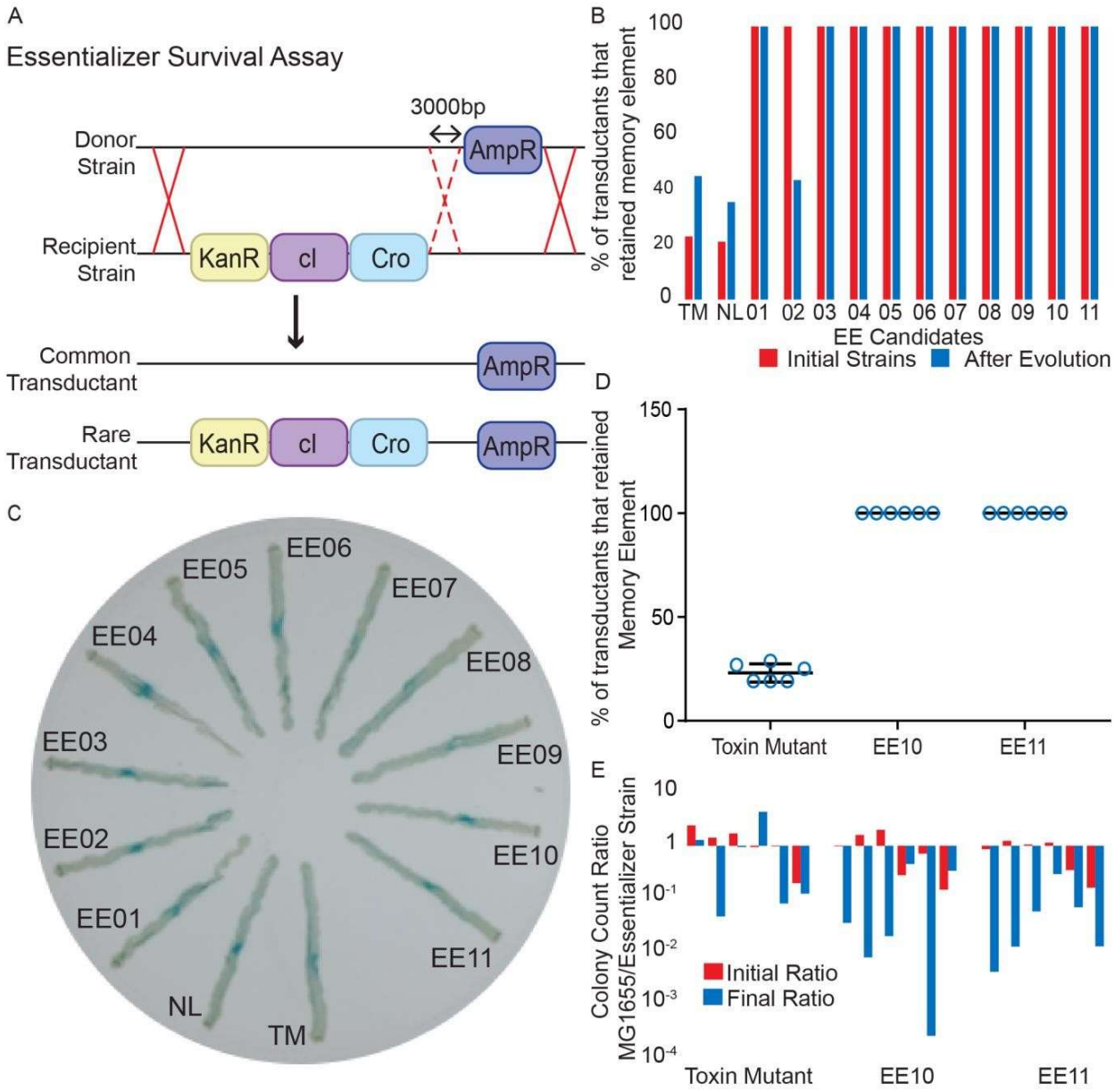

Figure 3: Analysis of essentializer element candidates: A) Transduction Assay. A donor strain with ampicillin resistance ~3000 bp from the memory element was used to remove the memory element. Due to the spacing between the loci of the cassettes a small subset of transductants would have both cassettes. **B)** Percentage of colonies that retained the memory element after the transduction assay (Figure 3A), both before (red) and after (blue) passaging for 140 generations. TM = toxin mutant and NL = EE non-lethal. **C)** Candidate essentializer strains were streaked on a plate spanning sub and super induction levels of ATc. Super induction of Cro near the center represses *lacZ* expression from the memory element. An intermediate expression level of Cro allows stable switching to the *cro* state and lacZ expression. Sub induction levels results in remaining in the cI state. **D)** Six biological repeats of the transduction assay (Figure 3A) for candidates EE10 and EE11.

E) Six biological repeats of a competitive growth assay conducted over 70 generations to compare the fitness of parental (MG1655) and engineered bacterial strains. See also Tables S1-3.

The lethal activity of essentializer element candidates remained stable over extended growth periods in the cI state (Figure 3B). Candidates were passaged for approximately 140 generations and then the test outlined in Figure 3A was used to assay whether the memory element was still essential, by screening 52 transductants for each candidate. Only candidate EE02, when subcultured, failed to maintain its selection for preservation of the memory element (Figure 3B). Sequencing of isolated clones from EE01-EE10 subcultures revealed multiple mutations in the regulatory region of both the toxin and the antitoxin for candidates EE02 and EE06, but no such mutations in the other candidates. It appears that the mutations in the regulatory region of EE06 alters the expression rate of the toxin and antitoxin but does not prevent the lethal effect of the essentializer. This result might be expected if there is a selection for some level of toxin expression in the presence of the antitoxin. The nature of the essentializer limits the screening for potential escapees to the transduction-based assay (Figure 3A). However, it can be inferred that had an essentializer mutation arisen within the population of candidates EE1, EE3-5 and EE7-10, such mutations were not able to confer a growth advantage and become a significant proportion of the population within the 140 generation time frame. During 140 cell divisions, it is expected that at least one adaptive sweep may occur (Maddamsetti et al., 2015, Novick & Szilard, 1950).

Upon switching to the Cro state of the memory element, candidates maintained viability with the presence of Cro rather than cI. PAS132 contains a trigger element, allowing for the expression of *cro* under a tetracycline sensitive promoter P_{tet} (Jonathan W Kotula et al., 2014). If grown in the presence of anhydrotetracycline (ATc), the memory element switches from the cI state to the Cro state through expression of the P_{tet} *cro* trigger. All candidates were streaked on a gradient of ATc spanning sub- induction to super-induction levels. At high ATc concentrations, Cro binds to the O_R1 and O_R2

operator regions in addition to its preferred O_R3, repressing itself in the memory element design. For this reason, switching to the Cro state can only occur in a narrow range of Cro concentrations. In Figure 3C we see narrow strips of blue for each candidate, indicating the strain has survived the memory element flipping from the cI state to the Cro state, and subsequently allowed expression of the reporter beta-galactosidase (LacZ).

The essentializer element candidates EE10 and EE11 were selected for more in depth assays of their evolutionary stability. Each candidate was passaged for approximately 140 generations, along with the toxin-defective control. The candidates were subsequently tested for transductional removal of the memory element (Figure 3A). For both EE10 and EE11, 52 of 52 colonies screened for each biological repeat maintained the memory element, compared with about 20% for the toxin mutant control (Figure 3D). In addition, co-cultures of roughly equal starting ratio of the essentializer candidates and the parental MG1655 strain were grown and passaged for approximately 70 generations. The ratio of essentializer strain to parental MG1655 was measured by comparing colony- forming units before and after growth across 6 biological repeats (Figure 3E). Strains EE10 and EE11 generally outgrew MG1655, implying that the essentializer element does not confer a significant growth disadvantage, and suggests that expression of certain ratios of toxin plus antitoxin could confer a selective advantage.

The sequences of the artificial regulatory regions were determined in essentializer elements EE01-EE11, as well as for several elements that failed at various points during the screening process. Overall, about 2/3 of the sequences had deletions of 1 or more bases in this region (Table S2). The remainder of *ccdA* and *ccdB* were not sequenced, and it is possible that additional mutations could have occurred in these genes or in the host genome. In five cases (the lethal candidates EE01 and EE07 as well as the 3 non-lethal candidates EE13, EE15 and EE27) the 5'-most operator (O_R2) had a deletion and yet the strains are viable to different extents, indicating that loss of this operator is not required for tight repression and suggesting that the system is somewhat overdesigned. However, it is difficult to

define a structure-activity relationship between the regulatory region sequences and the phenotypes observed, possibly because the widely varying ribosome binding sites may lead to changes in mRNA structure.

To test whether an antitoxin is valuable in these constructions, we designed versions of EE10 and EE11 with the antitoxin ORF removed. These constructs were genomically inserted into TB10 by recombineering. Eight and seven (respectively) recombinant colonies of the toxin-only EE10 and EE11 strains were picked and grown at 42°C to determine if they could survive without the repression from cI^{857}. Two strains that did select for the presence of the memory element were passaged for approximately 140 generations, before being subjected to the transduction assay (Figure 3A). All biological repeats for both constructs retained their selection for the memory element after this period of growth (Figure S5).

Design and Construction of the Temperature Sensitive Kill Switch 'Cryodeath'

To further test our approach to balancing toxin/ antitoxin expression, we designed a kill switch that would respond to temperature. This would allow confinement of an engineered bacterium to an environment with a defined temperature, such as the mammalian gut. The temperature sensitive regulatory region of cold shock protein A (P_{cspA}) is modular and confers temperature-regulated expression of a protein of interest (S. J. Lee et al., 1994). P_{cspA} contains a constitutive promoter that has a high rate of transcription at all temperatures (Yamanaka, 1999). This is followed by a long 5' untranslated region (UTR) of 159 bp that adopts an unstable secondary structure at 37 °C and is rapidly degraded by RNaseE (Mitta et al., 1997), but at lower temperatures forms a stable configuration that allows translation (Fang et al., 1997, Giuliodori et al., 2010). The cold-shock regulatory region also includes a downstream box (DB) element located within the first 13 amino acids of the open reading

frame of *cspA* that enhances translation during cold shock by binding to the anti-DB sequence of 16S rRNA (Etchegaray & Inouye, 1999, Mitta et al., 1997) (Figure 4A).

We demonstrated the utility of this isolated regulatory region by inducing GFP in a temperature sensitive fashion. As it was unclear whether a particular gene of interest would be affected by an additional 13 amino acids on the amino end, we generated two versions of the regulatory region fused to GFP - one with a linker (N-GGGGS-C) between the truncated CspA and GFP designed to minimize interaction of the DB element, and one with no linker (Figure 4A). When assayed for expression at 37 °C, no discernible expression was observed compared to negative controls. However, to a small extent at 30 °C , and greater extent at room temperature (referred to as 22 °C) and 15 °C, the level of induction was visibly increased both with and without the linker (Figure 4B and S2), suggesting that the regulatory region functioned in a temperature sensitive manner.

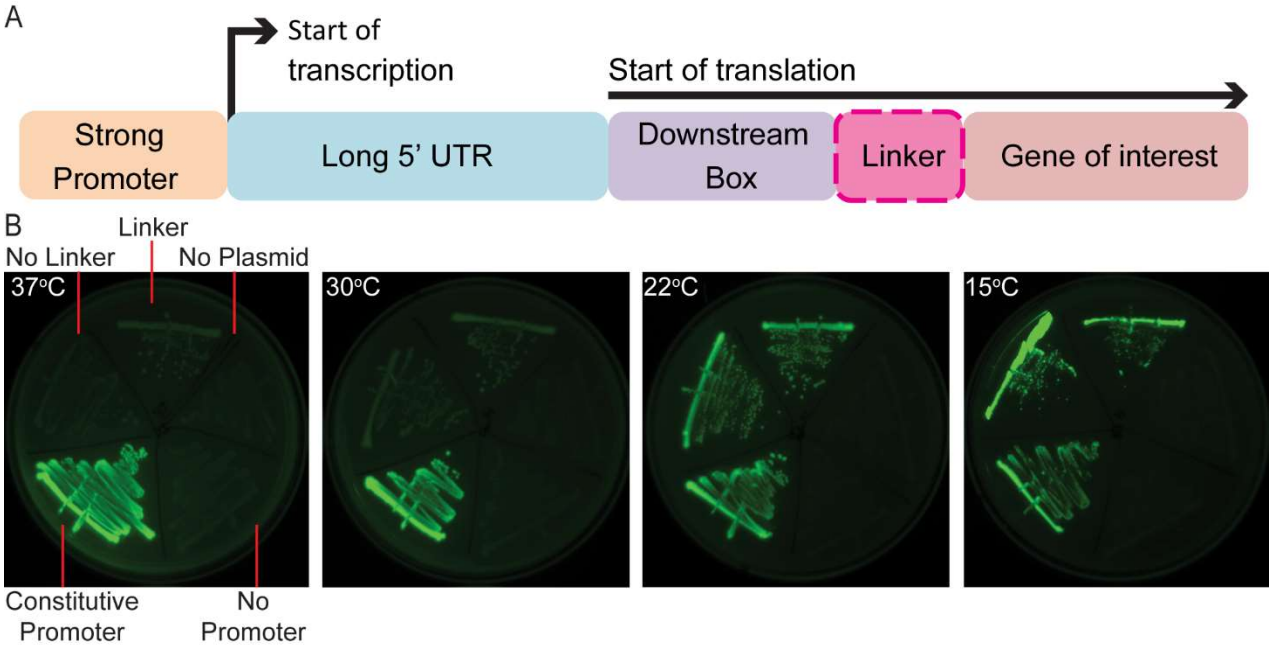

Figure 4: Structure and modularity of P*cspA* **A)** The regulatory region of cold shock protein A (CspA). **B)** Expression of GFP under P*cspA* at 37 °C , 30 °C, 22 °C, and 15 °C for 10 days. GFP expression level is shown with and without a linker, under P_{rpsL}, with no promoter and with no plasmid present. Images are from the same picture. See also Figure S2.

Rational design was used to generate small libraries to introduce variation at key locations in the regulatory region of the toxin and antitoxin. The *ccdB* coding region was placed after the *cspA* regulatory region, both with and without a linker. In a similar design to the essentializer element, the *ccdA* coding region was placed after a modified, constitutive LacUV5 promoter (P_{lac*}) (Malan & McClure, 1984) (Figures 5A and 5B). A Gentamycin resistance cassette was used for selection purposes. To achieve a range of expression levels for the toxin and antitoxin, ten bases were varied - three in the RBS of the antitoxin, two in the -10 promoter region of the antitoxin, one in the -35 promoter region of the toxin, two in the -10 promoter region of the toxin, and two in the RBS of the toxin (Figure 5C). Each varied position had the possibility of 2 different nucleotides, making a possible 2^{10} = 1024 different constructs. The bases chosen to be varied followed a similar logic to that of the essentializer element, generally with bases of less importance from within the -10 and -35 regions, according to previous published analysis of *E. coli* promoter and ribosome binding sites (Mulligan et al., 1984; Shultzaberger et al., 2001, Shultzaberger et al., 2007).

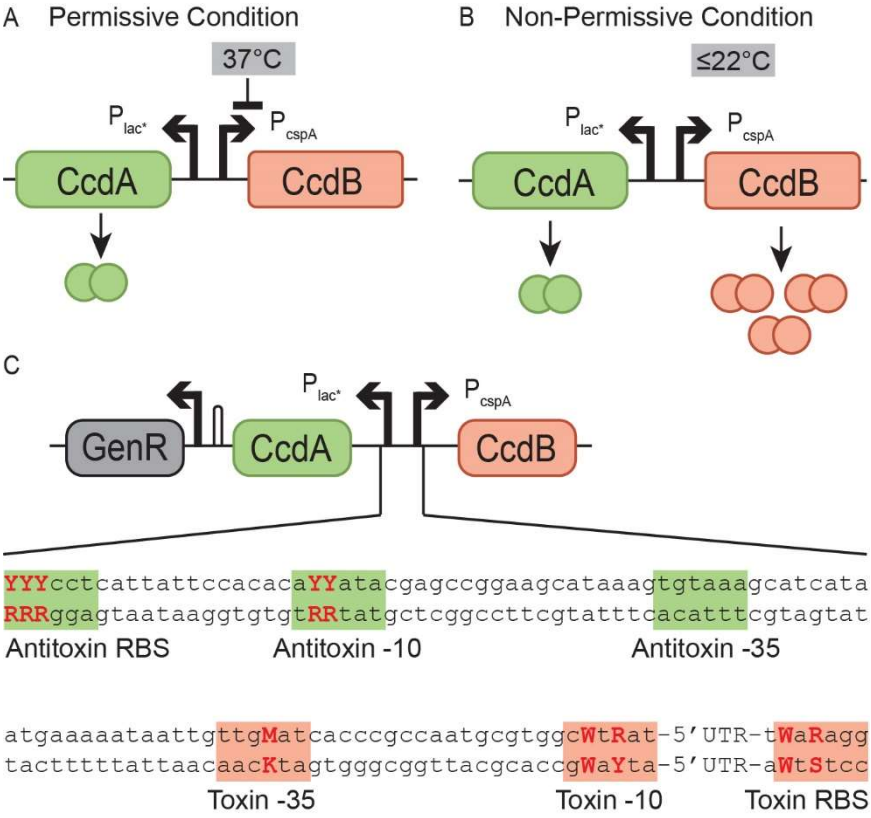

Figure 5: Design of the temperature sensitive kill switch cryodeath A) At 37 ^{O}C translation of CcdB is limited allowing cell survival. **B)** At colder temperatures, expression of CcdB increases, resulting in cell death **C)** The engineered region encompassing the regulatory DNA for CcdB and CcdA. Highlighted are segments that have a strong impact on expression level for both the toxin (red) and antitoxin (green). Varied bases are emphasized. Y = C or T; M = A or C; W = A or T; R = A or G. See also Figure S3.

To screen for cold-sensitive toxin induced death**,** *E. coli* strain Dh10β was transformed with the linear fragment libraries, using the lambda *red* genes expressed on plasmid pKD46 to enhance recombination. After an initial selection for gentamicin resistance, 26 unique candidates were identified; 2 with no linker and 24 with a linker. The 26 candidates were colony-purified on LB agar and tested on plates at the non-permissible temperatures of 30 ^{o}C , 22 ^{o}C and 15 ^{o}C as well as the permissible temperature, 37 ^{o}C. Of the 26 candidates, ten failed to grow or grew poorly at <37 ^{o}C, and were selected for further analysis (CD01-CD10). Non-lethal candidates were termed CD11-CD26. All except CD01 and CD11 contained a linker. As lethal candidates were identified both with and without a linker, the linker has no or little effect on the function of CcdB. CD12 was subsequently used as a negative control, referred to in this text as “CD non-lethal”. The candidates were sequenced across the antitoxin, toxin, and regulatory region, toxin and antitoxin open reading frames (Table S5). Although most candidates had retained the expected sequence across the regulatory region, only varying at the intended positions, candidate CD10 had a 9 bp deletion partially including and upstream of the toxin RBS (Figure S3). Modifying bases in this region has previously been noted to affect the stability of the CspA mRNA (Fang et al., 1997).

Evolutionarily stable circuits with the intended phenotype were identified. A survival assay comparing cfu at room temperature to that at 37 ^{o}C was used to quantify the extent of population death, termed the survival ratio (Figure 6A). After an initial assay was made at room temperature, the candidates were passaged in permissible conditions for approximately 140 generations in LB before a second assay for cold-sensitivity was performed at room temperature. The pre- and post-survival ratios for each candidate were compared to determine which candidates induced the highest level of

population death, and which candidates maintained their lethality after the opportunity for evolution (Figure 6B). CD01 and CD10 had low survival ratios of approximately 10^{-4} both before and after the period of growth, and were therefore chosen for further analysis. All candidates were sequenced across the antitoxin, toxin and regulatory region after the period of growth, with no mutations. For the candidates that have lost lethality, this implies that a mutation has occurred in another location in the genome to prevent the effect of CcdB.

To identify what these modes of escape could be, we used whole genome sequencing to identify mutations that arose in the cryodeath strain CD08 after 140 generations. Two colonies were identified from separate passaging repeats, and were assayed to ensure they were not cold sensitive (Figure 6A). These were sequenced and compared to the genome of CD08 before the period of growth. Of the two strains, one had no noticeable alterations between the parent and evolved strain, and the second had a singlxe SNP in the 5' UTR of the native *cspA* gene, as well as an approximately 500 bp deletion in the ORF of *gtrS* (Table S6). It is not immediately apparent what effect these mutations would have.

The cryodeath kill switch candidates were transferred to a different genetic background for use in the mammalian gut. After transfer by P1 transduction from DH10β into MG1655, a survival assay was used to measure lethality in both strains at 30 ^{o}C, 22 ^{o}C and 15 ^{o}C. In DH10β, both CD01 and CD10 induced no death compared to the toxin mutant control at 30 ^{o}C, a survival ratio ranging from 10^{-4} to 10^{-5} at room temperature and 10^{-5} at 15 ^{o}C (Figure S4). In MG1655, survival ratios were over an order of magnitude lower, 10^{-5} to 10^{-6} at 22 ^{o}C, and 10^{-6} for candidate CD10 at 15 ^{o}C (Figure 6C).

Candidate CD01 also had a substantial drop in survival ratio at 30 °C, (10^{-3}).

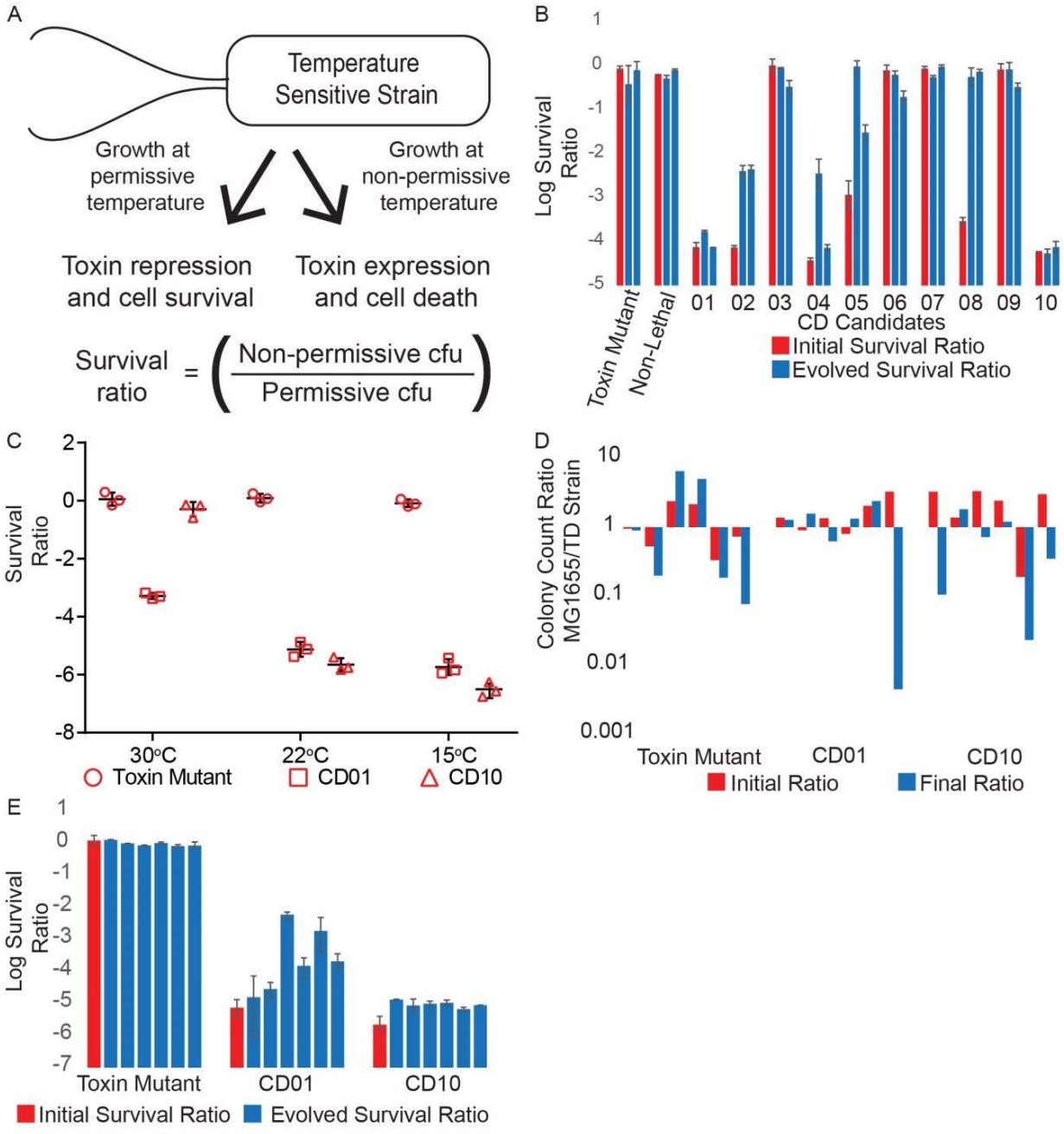

Figure 6. Analysis of cryodeath cadidates. A) Survival assay used to test the extent of population termination at non-permissible temperatures. **B)** Survival ratio of the ten temperature sensitive candidates in DH10β (CD1-CD10) before and after 140 generations of growth at a permissible temperature. Data represents average of two technical repeats, with error bars showing range. Two biological repeats after a period of growth are shown. Non-Lethal = CD non-lethal. **C)** Survival ratio of candidates CD01 and CD10 in MG1655 at 30 ^{O}C, 22 ^{O}C, and 15 ^{O}C. Three technical repeats of each condition are shown. **D)** Six biological repeats of a competitive growth assay of 70 generations to

compare the fitness of parental (MG1655) and engineered bacterial strains. **E)** Six biological replicates of survival ratio of candidates CD01 and CD10 in MG1655 after 140 generations, each biological replicate data point is the average of 3 technical replicates plotted with 3 technical repeats. Initial survival ratios are from the data collected for Figure 6C. See also Figure S4 and Table S5.

The cryodeath candidates showed no disadvantage in growth rate when compared to wild-type MG1655. Co-cultures were grown in minimal media for approximately 70 generations. The population ratio of kill switch strain to MG1655 was measured by comparing cfu before and after growth across 6 biological repeats (Figure 6D). Across the repeats, it varied which strain gained an evolutionary advantage over the time period, implying that a spontaneous mutation independent of the kill switch was responsible for the advantage.

One cryodeath strain was stable after an extensive period of growth. An evolutionary stability experiment was repeated with 6 biological repeats of CD01, CD10 and the toxin mutant control. Again the candidates were grown for approximately 140 generations and then assayed at room temperature to determine their evolved survival ratio. CD10 maintained its survival ratio at around 10^{-5} after the period of growth across all 6 biological repeats. The survival ratio for CD01 decreased by at least an order of magnitude in four cultures (Figure 6E) suggesting that a certain proportion of the population had lost or modified the lethal phenotype. As the candidate displaying the most robust evolutionary stability, CD10 was chosen as the most desirable cryodeath construct. To test if the antitoxin was necessary to maintain stability for CD10, we attempted to construct a version with no antitoxin. However, any such attempt resulted in a frame shift appearing in the *ccdB* ORF and a non-cold sensitive strain (Figure S5 and Table S8).

To determine the escape modes of CD10, we isolated 5 different escapees from the survival assay (Figure 6A), conducted at 22 ^{o}C. Their genomes were sequenced and assayed for differences to the parent strain, CD10. Of the 5, one had a frame shift mutation in *yedN*, one had a large insertion in the 5' UTR of *ccdB*, one had a 9 bp insertion between the -10 and -35 of the *ccdB* promoter and 2 had

no discernable differences when compared to CD10 (Table S6). Except for the mutation in *yedN*, it is easily deduced how these mutations would have an effect on toxin expression at 22 °C. We have no definitive explanation for the strains seemingly lacking a genomic mutation (although an unstable tandem duplication (Roth, 1978) would be difficult to detect by whole-genome sequencing). However, we note that in this selection we isolated promoter mutants that may have partial activity, while when selecting in the absence of the antitoxin, we isolated mutations in the toxin coding sequence. These observations suggest that expression of the antitoxin in the absence of the toxin may be somewhat deleterious.

CD10 induced efficient and stable population death upon defecation from the mammalian gut. Streptomycin resistance was introduced to CD10 and the toxin mutant control in order to facilitate colonization of the mouse gut by P1 transducing a mutated *rpsL* from a spontaneously streptomycin resistant strain. CD10 and the toxin mutant negative control were each gavaged into 3 separate Balb C mice after 24 hours of streptomycin treatment, and fecal samples were collected before gavaging, 24 hours after gavaging and 10 days after gavaging. The survival assay (Figure 6A) was then carried out on cultures grown from these fecal samples. When compared to earlier in vitro experiments CD10 samples showed a similar level of lethality with a survival ratio of 10^{-5} across all mice and time points (Figure 7B). No bacteria were isolated on streptomycin plates from the sample collected before gavaging.

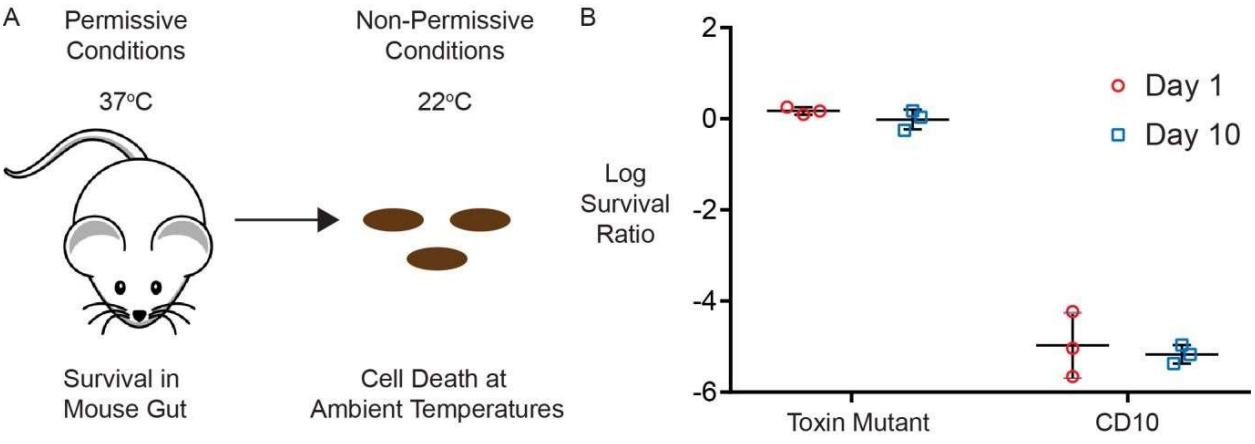

Figure 7: Testing cryodeath in vivo. A) Containment of transgenic bacteria to mammalian gut by temperature sensitive induction of cryodeath. **B)** Survival assay of cultures grown from feces of 3 mice gavaged with toxin mutant and three mice gavaged with CD10 in MG1655, both 1 day and 10 days after gavaging. Each data point is a separate biological repeat, made from the average of three technical repeats.

3.4 Discussion

We have constructed two inducible toxin/antitoxin kill switch systems for controlling the environment or conditions in which a genetically engineered strain of *E. coli* can survive (Figure 1). The first is the "essentializer" element, a kill switch that links cell survival to the presence of the memory element, an engineered genetic circuit that is normally not essential (Figures 2 and 3). The second, "cryodeath," responds to environmental temperature, allowing growth at 37 ^{o}C but resulting in a survival ratio of less than 10^{-5} at 22 ^{o}C and below (Figures 5, 6, and 7). Using these two systems, it is possible to limit the growth conditions of transgenic *E. coli* without the need for human monitoring or input.

The design criteria for these kill switches involve multiple phenotypes, including a lack of detrimental effect on viability and growth, which collectively do not lend themselves to high-throughput screening. Our approached therefore relied on rational design and construction of small libraries whose members could be extensively tested. Promoter and ribosome binding site expression levels were varied to create a pool of candidates with different levels of toxin and antitoxin production. Overlap extension PCR or Gibson assembly with degenerate primers were used to vary primarily those bases expected to modulate the quantitative level of gene expression but not affect regulation. This method was chosen over random mutagenesis technique such as error prone PCR, as the limited pool of potential candidates could be screened in its entirety if needed, and because an intense random mutagenesis would likely create primarily loss-of-function mutations that could mask mutations of interest. In addition, as many of the possible nucleotides have a well characterized effect on expression

(Mulligan et al., 1984, Shultzaberger et al., 2001, Shultzaberger et al., 2007), we could limit the potential candidates to those that were more likely to perform within the desired range of expression. Another alternative approach, use of wild- type promoters with varying strengths, was not chosen because such promoters may be regulated in an unknown manner, and because the chosen approach was technically simpler, faster, and less expensive. The success of modifying two different regulatory regions indicates this technique could be applied to any regulatory region where the influence of each individual base can be identified with a reasonable degree of accuracy, i.e. any promoter in *E. coli* or other well characterized organisms. Due to the error prone nature of oligo synthesis, many of the constructs sequenced had mutations likely to affect regulatory function, including the most stable cryodeath candidate CD10.

Both kill switches maintain functionality after at least 140 generations (20 passages) *in vitro* (Figure 3 and 6). The cryodeath CD10 element maintained 100% of its functionality over this time (Figure 6). In addition, CD10 maintained its lethality when used *in vivo* to limit bacterial growth after expulsion from the mammalian digestive tract in a mouse model. Under these conditions, cryodeath remained functional for at least 10 days (Figure 7). Of the 10 cryodeath strains that originally showed a temperature sensitivity at 22 ^{o}C, only 1 proved to be stable after an opportunity to evolve as opposed to 10 out of 11 essentializer candidates. This can likely be attributed to two causes; a lower basal expression level from P_R compared to P_{cspA}, and the greater opportunity for the unstable essentializer candidates to lose their lethality throughout the more extensive screening process.

These observations indicate that the kill switches designed here are likely to be genetically stable. In this context, it is important to note that the phenomenon of "periodic selection" has a conserving effect on evolution in asexual populations (Atwood, 1951; Novick & Szilard, 1950). Essentially, for any given environment, favorable mutations will arise in such a population and the

mutants will overtake the parental population. Based on inspection of data from long-term cultures (Maddamsetti et al., 2015, Novick & Szilard, 1950), a typical selection coefficient for such an overtaking mutation is about 1-10%. If a kill switch is negatively selected with a coefficient of 0.5%, then in a long-term culture of bacteria containing a kill switch, there will arise unlinked mutations with a selection coefficient of at least 1% and mutations in the kill switch with a selection coefficient of 0.5%; bacteria with the unlinked mutation, which will in general have a non-mutant kill switch, will overtake the population and thus the kill switch will be preserved even though it is slightly deleterious.

Our working hypothesis is that the presence of the antitoxin increases the stability of the system by negating the deleterious evolutionary pressure of leaky toxin expression. To test this, we designed constructs identical to our best performing kill switches (EE10, EE11 and CD10) except with the antitoxin ORF removed. The CD10 homolog failed to yield a stable cold-sensitive strain (Figure S5 and Table S8) implying that basal level of toxin expression at permissible temperatures prevents *E. coli* survival without the antitoxin. For the essentializer homologs, while constructing a strain with the memory element and a toxin-only cassette over 25% of recombinants tested unable to select for the presence of a cI repressor (Table S4). However, once the toxin only cassettes were combined with the memory element, they were stable over an equal time length to that tested for EE10 and EE11. It is likely that cI^{857} has a lower strength of repression than its wild-type counterpart (Angeles, 1976) and whilst the antitoxin does not appear necessary for EE10 and EE11 to maintain their functionality, it was likely beneficial during construction when repression is less than optimal.

The cryodeath system consistently reported a survival ratio of at least 10^{-5} at 22 ^{o}C (Figure 6 and 7) and can reach as low a survival ratio as 10^{-6} at 15 ^{o}C (Figure 6C). P*cspA* has been shown to induce a 16 fold increase in expression as high as 27 ^{o}C, indicating the potential for a lethal level of induction at higher temperatures (Hoynes-O'Connor 2017). Previously published kill switches using the toxin CcdB as the sole source of cell death in *E. coli* have a survival ratio of approximately 10^{-3} (

(Chan et al., 2015; Piraner et al., 2016). This disparity in escape frequency can be partially explained by the fact that the target of CcdB, the GyrA subunit of DNA gyrase, is one of the genes upregulated by the cold shock response (Jones 1992). In addition, the fact that our approach yielded a kill switch that induced a higher frequency of population death can be attributed to the screening of a rationally designed small library.

An important design requirement is that kill switches should not be deleterious to growth of the host organism in permissible conditions. This means that stochastic variation in expression of the toxin and antitoxin genes should not allow greater expression of toxin even in rare circumstances. Natural toxin-antitoxin systems avoid spontaneous deletion because the antitoxin is much less stable than the toxin, so a stochastic decrease in antitoxin expression will not be significantly averaged over time. The antitoxin promoter used here is based on the *lac* promoter, which initiates transcription about once per minute when induced (Bremer, 1975), corresponding to an average of 20 events per generation in rapidly growing cells. If the rate of transcription of the toxin promoter is averaged at roughly one quarter of this (5 events per generation), there is less than a 1% chance that the rate of toxin transcription will be greater than the rate of antitoxin transcription at any given time. However, only a fourfold increase in toxin transcription increases the likelihood of an excess of toxin expression to greater than 50% (Figure S6).

A variety of promoters responding to different environmental conditions have been identified in *E. coli*, and our system could be modified to include an alternative mode of death induced by pH (Chou, 1995), oxygen level (Cotte, 1990), or nutrient abundance (Yansura & Henner, 1990). Combined with an independent toxin-antitoxin system (Yamaguchi & Inouye, 2011), the escape frequency of the system would be decreased whilst maintaining a high level of evolutionary stability.

Contributions:

Conceptualization, F.S. and J.W.; Investigation, F.S., L.B., S.O., E.R., and J.O.; Writing – Original Draft, F.S.; Writing – Review & Editing, F.S., J.W., P.S., L.B., and J.O.; Funding Acquisition, P.S. and J.W.

Acknowledgements:

We thank Georg Gerber for help with building a theoretical model for kill switch design. We thank Sarah Boswell for guidance with preparing Miseq samples for whole genome sequencing. We thank Sean Wilson from the laboratory of Prof. Ethan Garner for help with image analysis.

Funding:

This work was supported by Defense Advanced Research Projects Agency Grant HR0011-15-C-0094 and funds from the Wyss Institute for Biologically Inspired Engineering. FS acknowledges funding from NIH training grant [5T32GM007598].

3.5 Materials and Methods

Contact for reagent and resource sharing

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Pamela Silver (pamela_silver@hms.harvard.edu)

Experimental model and subject details

***E. coli* K12 strain MG1655**

Used as the bases for all strains, except for the initial screen of cryodeath constructs that was conducted in DH10β. Maintained using established protocols for *E. coli*. Strains that contained cryodeath candidates were maintained at a constant temperature of 37 ^{o}C unless being assayed, or kept as glycerol stocks at -80 ^{o}C

***E. coli* K12 strain TB10**

A derivative of MG1655, with a large section of the lambda prophage genome inserted into biotin operon. Used to recombine the essentializer library. cI has been mutated to the temperature sensitive variant, cI^{857}, allowing for temperature sensitive induction of the lambda *red* genes. Maintained using established protocols for *E. coli*, except all growth was kept at 30 ^{o}C unless being assayed.

***E. coli* strain DH10β containing plasmid pKD46**

E. coli strain used to recombine cryodeath library. Maintained using established protocols for *E. coli*. Before cryodeath integration, strain was maintained at 30 ^{o}C to allow for pKD46 propogation. After cryodeath integration, strains were maintained at 37 ^{o}C unless being assayed, or kept as glycerol stocks at -80 ^{o}C.

Mouse Strain BalbC

Approval for animal work came from the Harvard Medical School IACUC under protocol 04966. Seven week old female BalbC mice from Charles River Laboratories were given two weeks to acclimatize to the facility upon delivery. Three mice were used to test CD10 and three used for the negative control, kept in two respective cages. Maintained using the standard conditions of the facility.

Method Details

Media and growth conditions:

Unless otherwise specified LB media with relevant antibiotic was use for all growth conditions. For the competitive growth assay, screening for lethality of the essentializer candidates and for the initial evolutionary screen of the essentializer candidates, M9 minimal media supplemented with 1 mM $MgSO_4$, 1 µg/ml thiamine hydrochloride, 0.4% w/v glucose, 100 µM$CaCl_2$ was used. For all evolution experiments and survival assays no antibiotic were used (except for streptomycin in the fecal sample survival assay). For mouse survival assays, MacConkey lactose with streptomycin was used. For plating the competitive growth assay, MacConkey rhamnose without antibiotic was used to differentiate between wild-type MG1655 and strains with kill switches inserted into the rhamnose operon. Unless otherwise stated, in all instances of antibiotic use the following concentrations were employed: Amp 100 µg/ml, Kan 50 µg/ml, Gent 10 µg/ml, Strep 100ug/ml.

Cold-sensitive reporter plasmid design:

The P_{cspA} regulatory region was amplified using PCR from the K12 genome using primers TS3 and TS4 or TS6 to add homology to the plasmid pUA66, and incorporate the GGGGS linker in the case of TS6. pUA66 GFP was linearized by PCR using primers TS1 and TS2 and combined with P_{cspA} promoters

using a NEB Gibson assembly kit according to the manufacturers guidelines, to create plasmids with temperature sensitive expression of GFP.

Construction of kill switch libraries:

All primer sequences are given in Table S7. To construct essentializer variants (Figure 2D) a cassette was ordered as a Gblock from Integrated DNA Technologies with the sequence "original EE sequence" (Table S1). The degenerate oligo FS1 as well as primers FS2, FS3 and FS4 were used to amplify the cassette in two parts, which were subsequently combined using overlap extension PCR and amplified with FS3 and FS4. For the cryodeath kill switch, the degenerate primers TS12 and TS13 or TS18 were used to amplify P*cspA* with and without a linker respectively. Primers TS14-TS17 were used to amplify "original EE sequence" in two parts, which were subsequently combined with the degenerate P*cspA* amplicons using Gibson assembly and amplified using TS14 and TS17. In initial experiments, we found that the stitched product appeared to be correct based on gel electrophoresis, but failed to produce transformants upon recombineering (see below). We hypothesized that during the stitching amplification, the amount of DNA product exceeded the dNTPs and/or primers, with the result that the final PCR cycles simply involved denaturation and reannealing of full-length single strands, which would in general contain numerous mismatches. Upon transformation and recombination into the genome, such mismatches would be expected to undergo mismatch repair in a strand-independent manner, resulting in double-stranded breaks as gap-repairing polymerases meet. To avoid this problem, the reaction was passed through a Zymo Clean and Concentrator™ kit before being added to fresh PCR reagents for an additional cycle with primers FS3 and FS4 or TS14 and TS17 for the essentializer and cryodeath libraries respectively. This ensured that the predominant PCR product was a matching double stranded helix. After this modification, the frequency of transformant isolation increased dramatically.

P1 transduction:

P1 lysates conducted according to previously published methods (L. C. Thomason, Costantino, & Court, 2007) with lysate of greater than 10^9 pfu. Overnight cultures of donor strains were grown at relevant temperatures in LB, before being back diluted into 50 fold into 5 ml of LB supplemented with 0.2% w/v glucose, 20 mM $MgCl_2$ and 5 mM $CaCl_2$. After approximately 45 minutes of growth at the relevant temperature, between 10 and 100 ul of high titre (pfu > 10^9) lisate was added and the culture was grown until lysis had occurred. Lysates were then centrifuged at 4000 rpm for 5 minutes and the supernatant filtered using a 0.2 micron filter. Lysates were stored for up to 6 months at 4^{o}C.

Recombineering:

For both kill switches, electrocompetent cells were prepared for transformation using previously published methds (L. Thomason et al., 2007). The degenerate essentializer library was transformed into the recombinant strain TB10 that had been induced for 15 minutes in a 42^{o}C water bath with manual shaking every two minutes. For the cryodeath degenerate library, the recombinant plasmid pKD46 (Datsenko & Wanner, 2000) was transformed into Thermo Scientific MAX Efficiency DH10β electrocompetent cells. Cells were induced for 2-3 hours in the presence of 10 mM L-arabinose. Approximately 100 ng of DNA was combined with 50 µl of cells in a 0.1 cm cuvette, and electroporated using EC1 setting on a Biorad Micropulser. Cells were recovered in SOC medium for one hour before being spread on the relevant antibiotic plate.

Temperature sensitive survival assay:

All colonies from the recombination of the two temperature sensitive libraries (with and without a linker) were grown at 37 ^{O}C, 30 ^{O}C, 22 ^{O}C and 15 ^{O}C. 5ml overnight cultures were made at 37 ^{O}C of all samples from the 37 ^{O}C plate, and combined 1:1 with 50% glycerol at 37 ^{O}C before being immediately transferred to a -80^{O}C freezer. Unless being assayed for growth at lower temperatures, all subsequent experiments involving temperature sensitive strains were prepared in a constant 37 ^{O}C environment. For the strains that showed variation of growth at different temperatures, the glycerol stocks were used to make overnight cultures at 37 ^{O}C in 5ml of LB. This was diluted by a factor of 10 into 4 ml of LB in a 15 ml culture tube, and grown at 37 ^{O}C until ~OD_{600} 1.0 was reached. Serial dilutions of the cultures were made and plated onto a number of LB plates corresponding to the number of temperatures being assayed. Separate plates were then incubated for up to 15 days at various temperatures, and the cfu at each temperature compared to the cfu at 37 ^{O}C was termed the survival ratio.

Testing evolutionary stability of kill switch strains:

For the essentializer candidates, glycerol stocks were made of the 11 candidates that proved lethal from the extensive screening outlined in Figure 2E. These glycerol stocks were used to inoculate 2 ml of M9 minimal media without antibiotic. After 24 hours of growth, 20 µl was used to inoculate 2 ml of M9 minimal media. This was repeated for 20 passages. A P1 transduction was then used to potentially replace the memory element with ampicillin resistance (Figure 3A). Essentially, an ampicillin-resistance cassette was placed in the *mhpC* ORF, such that in a typical transduction, the ampicillin resistance and kanamycin resistance markers are about 80% linked. For each candidate (EE01-EE10), 52 ampicillin-resistant transductants were streaked onto ampicillin/kanamycin plates to determine whether the

memory element was still present. If 100% of transductants retained the memory element, it was deemed that the essentializer was still producing a selective pressure to keep the memory element.

Competitive Growth Assay:

Overnight cultures of MG1655, Toxin Mutant, EE10, EE11, CD01 and CD10 were grown to stationary phase. 100ul of each culture was then used to inoculate M9 minimal media, and glycerol stocks were made of this initial inoculation. Cultures were passaged once every 24 hours diluting 1:1000 into M9 minimal media. After 8 days, glycerol stocks were made of the final culture. Glycerol stocks were then defrosted and plated directly onto MacConkey Rhamnose plates. As the two kill switches had been inserted into the Rhamnose operon, they could be distinguished as making white colonies on Rhamnose MacConkey compared to red colonies produced by MG1655. Cfu were used to determine the ratio of kill switch strain to MG1655.

Testing Temperature Sensitive Kill Switch in the Mammalian Gut:

Approval for animal work came from the Harvard Medical School IACUC under protocol 04966. Seven week old female BalbC mice from Charles River Laboratories were given two weeks to acclimatize to the facility upon delivery. 500 ug/ml of streptomycin was added to their drinking water and maintained throughout the experiment. The Toxin mutant and strain CD10 were gavaged at a concentration of approximately 10^8 cfu/ml, each into 3 separate mice. Feces was collected between 2-6pm from day 0 (immediately before gavaging) to day 10, and immediately placed on dry ice before being transferred to a -80 ^{o}C freezer. Feces was resuspended in PBS by shaking at 4 ^{o}C for 1 hr, then 100 µl was used to inoculate 5ml of MacConkey lactose broth with streptomycin and grown overnight. The following day this overnight culture was diluted by a factor of 10 into MacConkey lactose and grown to

approximately OD 1.0. Cultures were then serial diluted and plated at 37 ^{O}C and room temperature to compare cfu at permissible and non-permissible temperatures.

Quantification and statistical analysis

For all graphs, error bars represent the range of data collected. For survival assays, data points were omitted if contamination was observed, or if the cfu from several different dilutions was equivalent, indicating a range of values for a particular cultures cfu spanning multiple orders of magnitude. In such cases a dilution error is suspected, and alternative technical repeats were used instead.

Chapter 4 – Synthetic circuits for pH mediated sensing, counting and containment

Finn Stirling*[1,2], Alexander Naydich*[1,2], Juliet Bramante[1], Rachel Barocio[1], Michael Certo[1,2], Hannah Wellington[1], Elizabeth Redfield[1], Samuel O'Keefe[1], Sherry Gao[1], Adam Cusolito[1], Jeffrey Way[1,2], Pamela Silver[1,2†]

[1]Department of Systems Biology, Harvard Medical School, 200 Longwood Avenue, Warren Alpert 536, Boston, MA 02115, USA

[2]Wyss Institute for Biologically Inspired Engineering, Harvard University, 3 Blackfan Circle, 5th Floor, Boston, MA 02115, USA

*These authors contributed equally to this work

[†]Correspondence: pamela_silver@hms.harvard.edu

This article is currently under review at Cell Reports

Published online on Bioarchives:
doi: 10.1101/740902

My Contribution

All work related to Figure 1 and Figure 2 was carried out by me or under my direct supervision by one of the students working with me; Juliet Bramante, Rachel Barocio, Hannah Wellington, Elizabeth Redfield, Samuel O'Keefe, Sherry Gao and Adam Cusolito. I was not involved with experimental design or implementation for work related to Figure 3 which was predominantly carried out by Alexander Naydich, based on Michael Certo's initial designs. For Figures 4 and 5, experimental design was carried out by me and Alexander Naydich, and data collection was carried out predominantly by me with assistance from Alexander Naydich. For the final manuscript, I wrote all parts except for the results, methods and discussion that relate to Figure 3, which were written by Alexander Naydich.

4.1 Abstract

As pH is fundamental to all biological processes, pH-responsive bacterial genetic circuits enable precise sensing in any environment. Where unintentional release of engineered bacteria poses a concern, coupling pH sensing to expression of a toxin creates an effective bacterial containment system. Here, we present a pH-sensitive kill switch (acidic Termination of Replicating Population; acidTRP), based on the *E. coli asr* promoter, with a survival ratio of less than 1 in 10^6. We integrate acidTRP with cryodeath to produce a two-factor containment system with a combined survival ratio of less than 1 in 10^{11} whilst maintaining evolutionary stability. We further develop a pulse-counting circuit with single cell readout for each administered stimulus pulse. We use this pulse-counter to record multiple pH changes and combine it with acidTRP to make a two-count acid-sensitive kill switch. These results demonstrate the ability to build complex genetic systems for biological containment.

4.2 Introduction

pH is an essential aspect of all biological processes. Sensing and responding to pH is therefore a powerful tool in the arsenal of a synthetic biologist, allowing for an engineered strain to be controlled by a variety of environments. For example, bacterial species that travel through the mammalian digestive system are exposed to multiple levels of acidity (Fallinborf, 1999; McConnell et al., 2007). Endocytosis of pathogens by the human immune system results in a decrease in environmental pH for the pathogenic bacteria (Geisow & Evans, 1984; Wang et al., 2017). The soil microbiome senses and responds to the fluctuations in pH (Rousk et al., 2010).

E. coli responds to a wide array of stressful situations, allowing it to thrive in a variety of conditions (Arsène et al., 2000; Babai & Ron, 1998; Baranyi et al., 2014; Bläsi & Young, 1996; Rodrigues & Rodrigues, 2018). In response to a rapid reduction of pH, the acid shock response gene *asr* is upregulated (Motieju, 2009; Šeputiene et al., 2004). The P_{asr} promoter is pH- responsive allowing pH-dependent control of the expression of a gene of interest (Hoynes-O'Connor et al., 2017).

A key aspect of bacterial control is the capacity to dictate what environment an engineered strain is capable of surviving in. For engineered bacterial strains to be deployed in clinical or environmental settings outside of a research laboratory, there must be a degree of confidence that they cannot escape to surrounding environments. Previously engineered forms of containment have used kill switches that rely on a small molecule survival factor to be provided by human intervention to ensure survival by i) allowing for expression of an essential gene (Cai et al., 2015); ii) regulation of a toxic gene (Caliando & Voigt, 2015; Contreras et al., 1991; Dirks et al., 2007; Ronchel & Ramos, 2001); iii) supplementation of an auxotrophy for a non-natural amino acid (Rovner et al., 2015), or a combination of multiple methods (Chan et al., 2015; Gallagher et al., 2015). Coupling containment to an environmental condition such as pH is an alternative way of controlling bacterial survival.

For any kill switch, the evolutionary stability of the system is paramount to its effectiveness. Often, a containment system designed to control bacterial growth causes a fitness defect, and hence a mutant defective in the killing function would have a growth advantage (Chan et al., 2015). If cell death is initiated by the up-regulation of a toxic protein, instability of a kill switch can be attributed to the basal expression of the toxic gene even in permissible conditions. Previously we and others have shown that the inclusion of an antitoxin, and careful balancing of toxin and antitoxin expression levels across relevant conditions, can mitigate this effect, resulting in a kill switch that is evolutionarily stable over biologically relevant time periods (Gallagher et al., 2015; F. Stirling et al., 2017).

A containment system must also have a survival ratio (defined as the fraction of the population able to survive in the non-permissible environment) low enough that the probability of an escape event is essentially zero. Previously published kill switches that respond to intrinsic characteristics of their environment have had survival ratios no lower than 10^{-6} (Piraner et al., 2016; Stirling et al., 2017), corresponding roughly to the mutation rate of a single gene (Lee et al., 2012). In practical environmental applications, population sizes are much greater than 10^6 cells, so even lower survival ratios are desirable. Here, we present a containment system with a survival ratio of less than 1 in 10^{11}.

A pulse counter is a specific type of recording circuit which tracks the number of discrete pulses of a single stimulus. Biological pulse counters might be used to monitor extracellular factors, such as fluctuations in antibiotic concentration at remote infection sites, or the rate of intracellular processes, such as bacterial division or metabolism. They may also be used to program a cellular response after a specific number of exposures. An effective cellular pulse counter demands three key characteristics: i) the circuit must exhibit a high signal-to-noise ratio, preventing advancement of the counter in the absence of a stimulus; ii) the circuit must produce a clear digital readout, such that the count can be interpreted on a single-cell basis; and iii) the circuit must advance the count only once for each discrete stimulus pulse, regardless of how long the pulse lasts. This last feature has been especially challenging

for previous synthetic biological counters, since in the absence of negative feedback, multiple counts may be recorded if the duration and concentration of the stimulus pulse are not carefully controlled (Friedland et al., 2009). Since the first attempts at constructing pulse counter circuits (Friedland et al., 2009), several mechanisms have been proposed for how a counter might avoid erroneously registering multiple counts for a single stimulus by delaying its response until the falling edge of a stimulus pulse, but none have been implemented to date (Noman et al., 2016; Subsoontorn & Endy, 2012). Here, we design a robust pulse counting circuit and combine it with our kill switch to demonstrate our ability to program cellular behaviour in response to a specific number of recurrences of a stimulus.

4.3 Results

Construction and Testing of the pH sensitive kill switch acidTRP

We constructed a pH-sensitive kill switch, Acid Termination of Replicating Population (or acidTRP), based on the *E. coli* P_{asr} promoter. P_{asr} is a pH-sensitive promoter native to *E. coli* that is repressed at pH 7 and induced at pH 5 (Figure 1A). It controls expression of *asr* (acid shock RNA), encoding a protein of unknown function that is heavily upregulated during the acid shock response (Šeputiene et al., 2004). Regulation of P_{asr} is under dual control of the *PhoBR* operon as well as the *RstBA* operon. Both PhoB and RstA binding are known to affect acid sensitivity (Baek & Lee, 2006; Ogasawara et al., 2007; Deliene, 1999).

In order to maintain stable control over a bacterial population, a containment system should not lose its function due to mutations over multiple generations. To achieve this, a toxin/antitoxin system was used, in which the antitoxin is constitutively expressed at a low level to mitigate any detrimental effect due to leaky expression of the toxin in permissible conditions. We used the type II toxin-antitoxin system Doc/Phd. Doc phosphorylates the elongation factor EF-Tu, inhibiting translation and preventing

growth (Castro-Roa et al., 2013; Liu et al., 2008). The antitoxin Phd binds to Doc, inhibiting the kinase activity (McKinley & Magnuson, 2005).

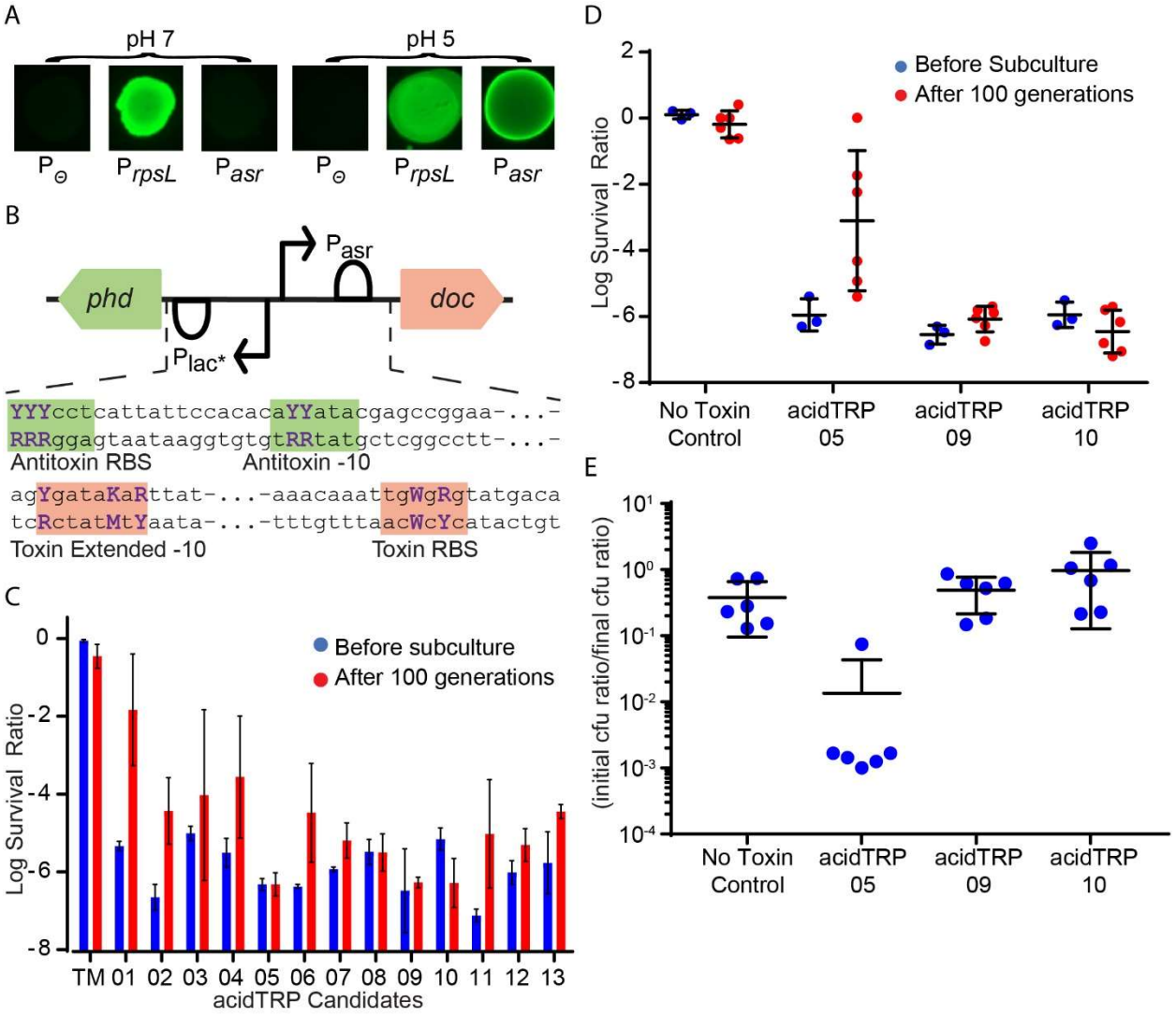

Figure 1. Construction and screening of a pH sensitive kill switch library. 1A) Expression of native P_{asr}-GFP at pH 7 vs pH 5. Left to right in each set: no promoter:GFP (negative control), P_{rpsl}:GFP (positive control), P_{asr}:GFP. Images are from the same picture, with comparable levels of bacterial growth. **1B)** Structure of acidTRP genetic circuit and degenerate base locations. Black arrows: promoter elements, semi circle: ribosome binding site, coloured arrows: genes. The sequence of the antitoxin and toxin regulatory regions are shown with the RNAP -10 binding sites and ribosome binding sites highlighted (antitoxin green, toxin red). Purple bases indicate degenerate locations (Y = C/T, K = G/T, R = A/G, W = A/T). P_{lac*} indicates a constitutive lac promoter. **1C)** Survival ratio for acidTRP candidates in *E. coli* TB10 before (blue lines) and after (red lines) 100 generations of growth. TM = toxin mutant control. Error bars represent the standard deviation from two biological replicates. **1D)** Survival ratio for candidates AT05, AT09 and AT10 in *E. coli* K-12 MG1655 before (blue) and after (red) 100 generations of growth. 3

technical replicates where used for the before data points and 6 biological replicates were used for the after data points. Error bars represent the standard deviation. **1E)** Competitive growth assay between acidTRP strains and *E. coli* K12 MG1655. Mixed cultures with approximately equal titers of an acidTRP strain and MG1655 were grown at pH 7 for around 70 generations. (Initial ratio of acidTRP candidate to MG1655)/(final ratio of acidTRP candidate to MG1655) is plotted for co-cultures of ATTM, AT05, AT09 and AT10 each with MG1655. For candidate 05, the bottom 5 data points represent the limit of detection as no acidTRP-05 colonies were observed after co-culture with MG1655. Data points are 6 biological replicates. Error bars are standard deviation.

To screen for strains with low escape rates and optimal evolutionary stability, a kill switch library with varying rates of toxin and antitoxin expression was created via rational design. Phd is expressed under a modified constitutive LacUV5 promoter (Malan & McClure, 1984), with Doc expression controlled by P_{asr}. In a similar technique to the creation of the temperature sensitive kill switch cryodeath (Stirling et al., 2017), degenerate bases were introduced at three locations in the antitoxin RBS and two locations in the antitoxin -10 RNA polymerase binding site, as well as at two locations in the toxin RBS and three locations in the toxin -10 RNA polymerase binding site (Figure 1B). Each of these ten modifications can result in one of two possible nucleotides, giving a library size of $2^{10} = 1024$ possible combinations. A construct was made with a frameshift mutation in the toxin ORF to act as a negative control and named acidTRP toxin mutant (ATTM).

The Kill Switch Library was screened to identify candidates that had a low survival ratio and did not lose their function across generations. The library was transformed into the *E. coli* strain TB10 (Caulcott & Rhodes, 1986; Johnson et al., 2004) using the lambda *red* recombineering genes (Poteete & Fenton, 2000). 200 Individual colonies were re-streaked on pH 7 and pH 5 plates, 41 of which displayed a notable growth defect at pH 5. These colonies were then grown in pH 7 liquid media and plated on M9 buffered to either pH 7 or pH 5. The CFU at each pH was compared to determine a survival ratio for these candidates, and those with a survival ratio of less than 1 in 10^5 were identified as potential candidates (Figure 1C, blue bars). These candidates were named acidTRP-01 through 13 (AT01-AT13). Next, to assess evolutionary stability, each candidate was passaged at pH 7 for over 100 generations. Candidates AT05, AT09 and AT10 all showed survival ratios of 10^{-6} or less after this growth period (Figure

1C, red bars) and were selected for further analysis. Each was transferred into the *E. coli* K12 strain MG1655 via P1 transduction, and the evolution experiment was repeated with more biological replicates (Figure 1D). Whilst candidates AT09 and AT10 showed no discernible difference before and after a period of growth, AT05 had a noticeable increase in survival ratio.

In permissible conditions, candidates AT09 and AT10 showed no disadvantage in growth rate when compared to MG1655. Candidates AT05, AT09 and AT10 were each co-cultured with MG1655 in pH 7 media and passaged for approximately 60 generations, the population ratio of each kill switch strain to MG1655 was measured by comparing the CFUs before and after growth across 6 biological repeats (Figure 1E). Candidates AT09 and AT10 showed no significant difference in the ratios of kill switch strain to MG1655 before and after the period of growth when compared to a co-culture of the toxin mutant control and MG1655. For candidate AT05, the kill switch strain was consistently out-competed by MG1655.

With no discernible difference between the survival ratio or evolutionary stability of AT09 and AT10, AT09 was chosen for further experiments. To characterise an induction curve, GFP was expressed under promoter P_{asr}-AT09. Cultures were grown to mid log phase at a range of pHs, and GFP fluorescence was measured using flow cytometry. Induction began at pH 6, with over 6000-fold increase in GFP production between pH 6 and pH 4.4 (Figure S1).

AT09 was combined with cryodeath (Stirling et al., 2017) resulting in a multiplicative effect on survival ratio in non-permissible conditions. Cryodeath is a temperature sensitive kill switch that uses the regulatory system from cold shock protein A (*cspA*) and the toxin antitoxin system CcdB/CcdA. At 37 ^{o}C the toxin is repressed, but at 22 ^{o}C an increase in expression of the toxin leads to population death with a survival ratio of around 10^{-5}. The combined strain, containing both cryodeath and acidTRP (CD-AT) was grown at 37 ^{o}C in pH 7 media and then plated on both pH 7 and pH 5 plates, which were incubated

at either 37 °C or 22 °C. Strains with a frame-shift mutation in one or both of the toxins were assayed in parallel as controls. Whilst the strain with individual kill switches behaved as expected, with survival ratios of 10^{-5} (only cryodeath functional) to 10^{-6} (just acidTRP functional), the double kill switch strain displayed a survival ratio that could not be detected as no colonies were formed when plating on pH 5 media and incubating at 22 °C (Figure 2A), even when 10^{11} cfu was plated.

CD-AT did not lose its capacity for containment after 100 generations of growth. CD-AT was passaged at 37 degrees for approximately 100 generations, along with the toxin mutant controls. Survival ratios for all strains were comparable to before the period of evolution (Figure 2B).

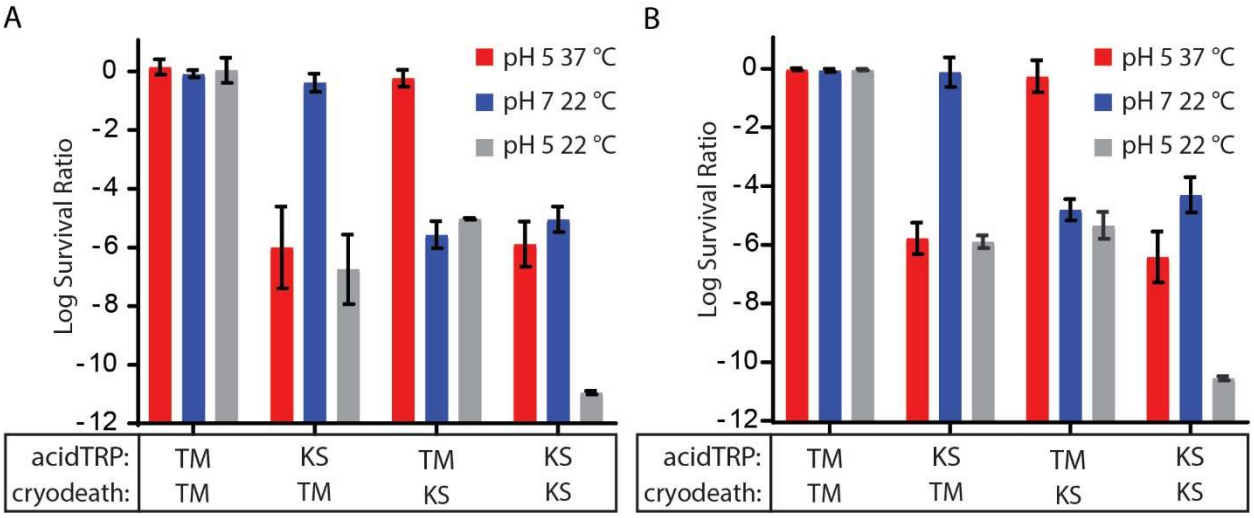

Figure 2. Combining the pH sensitive kill switch acidTRP and the temperature sensitive kill switch cryodeath (F. Stirling et al., 2017) **into a single strain. 2A)** The combination of acidTRP and cryodeath kill switches gives a multiplicative reduction in survival. Four strains with either a kill switch (KS) or toxin mutant (TM) inserted into the acidTRP and cryodeath loci were cultured in pH 7 media at 37 °C to late log phase. A dilution series was then plated on both pH 7 and pH 5 plates and grown at both 37 °C and 22 °C. Survival ratio for each strain was calculated when comparing the cfu at 37 °C/pH 5, 22 °C/ pH 7 and 22 °C/ pH 5 to the cfu at 37 °C/ pH 7. No colonies were observed on 22 °C/pH 5 plates for the double kill switch strain, so data points represent the limit of detection (ie total cfu on 37 °C/ pH 7 plate). Error bars represent the standard deviation from two technical replicates. **2B)** The combined kill switch strain is evolutionarily stable. The experiment from Figure 2A was repeated after a growth period of approximately 100 generations. Error bars represent the standard deviation from 6 biological repeats.

A novel counter circuit, responding to the falling edge of a stimulus pulse

We next constructed a counter circuit to allow our sensors to record multiple discrete pulses of a given stimulus. We built upon our previous design of a bacterial memory circuit, which is based on the OR operon of bacteriophage λ, and which is capable of sensing and recording transient stimuli (Jonathan W Kotula et al., 2014; Naydich et al., 2019; Riglar et al., 2017). We created a two-counter circuit (Figure 3A), based on the OR operon of bacteriophage 434, which produces a fluorescent response only after sensing two distinct pulses of a stimulus (e.g., low pH). Importantly, this circuit's ability to count pulses is not sensitive to an extended duration or concentration of the applied pulse. This feature is achieved by delaying the first count until the first stimulus pulse is removed (i.e., until the falling edge of the pulse).

The two-counter circuit consists of two constructs: a trigger, which senses a stimulus, and an actuator, which modifies the trigger after the first stimulus pulse is removed, thus enabling a different response to the second pulse. In our initial circuit, the trigger was designed to respond to anhydrotetracycline (aTc) using a P_{tet} promoter. The actuator is based on a bi-stable switch derived from the phage 434 OR operon, and its two states correspond to the mutually-repressive phage 434 proteins, Cro (under control of P_R) and cI (under control of P_{RM}) (Figure 3A). In our two-counter, the actuator begins in the cI state, with the P_{RM} promoter active. In this state, cI represses P_R, which controls the production of Cro and the λ Xis and Int proteins (Figure 3B, -1 step). When aTc is applied, Cro is produced from the trigger, while the terminator after the *cro* gene ensures that mCherry is not transcribed. The high level of Cro produced from the trigger in the presence of aTc places the actuator in a sustained intermediate state, with both P_{RM} and P_R repressed by Cro, due to the ability of Cro to repress P_R at high concentrations (Figure 3B, +1 step). When aTc is removed (the falling edge of the pulse), Cro is no longer produced from the trigger, and the level of Cro in the cell falls, leading to derepression of P_R. This, in turn, leads to expression of Cro from the P_R promoter, which maintains

repression of P_{RM}. It also leads to the production of Xis and Int, which excise the sequence between attL

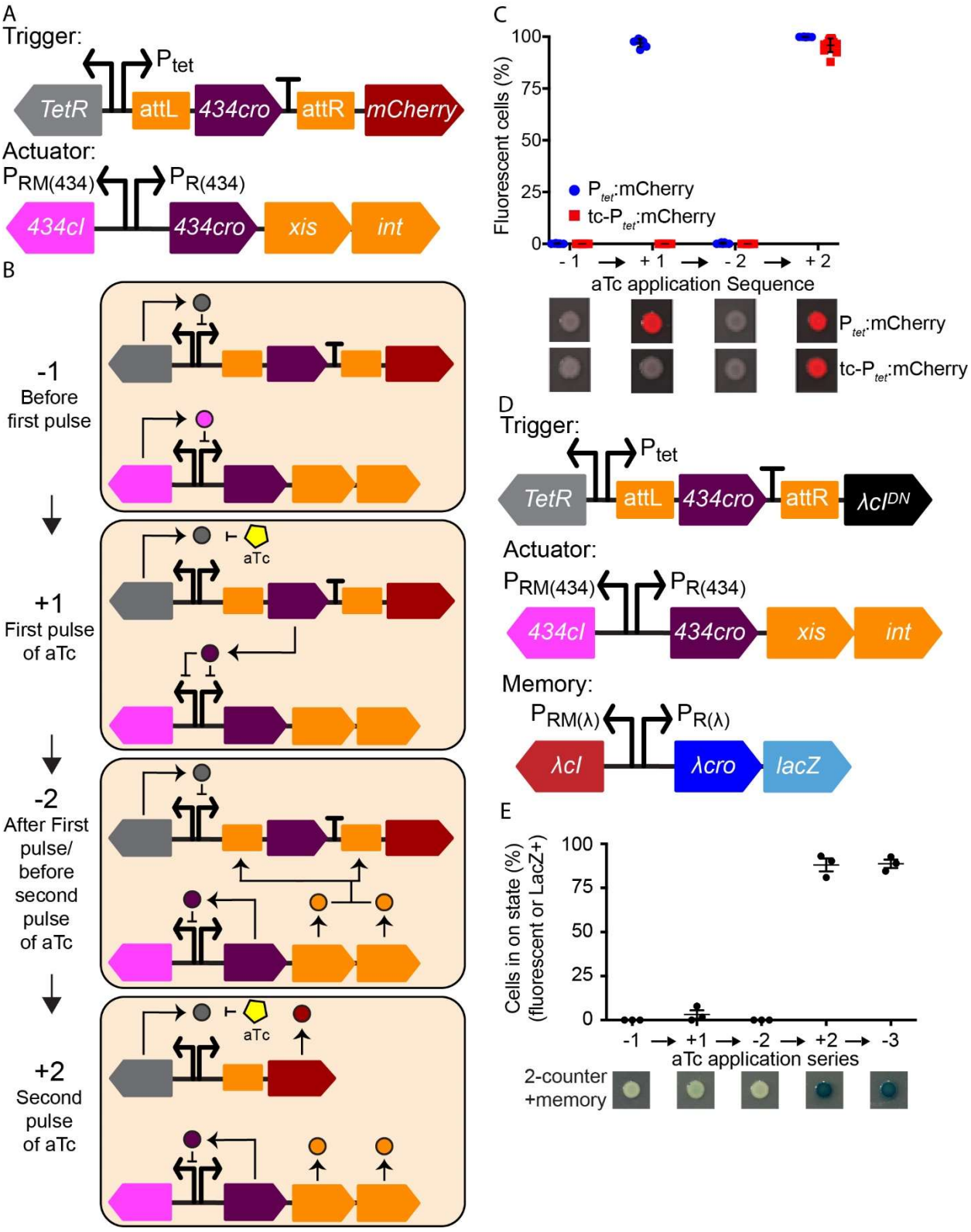

Figure 3: genetic circuits for 2 different two-counter circuits using an actuator based on phage 434 cI and Cro transcription factors. Black arrows: promoter elements, attL/attR: 434 Xis/Int attachment sites, T: terminator, coloured arrows: genes. **3A)** Schematic of the 434 two-counter, including trigger and actuator elements. **3B)** Schematic of trigger response and modification during two subsequent pulses of aTc. -1: actuator in cI state, TetR represses expression on the trigger. +1: aTc represses TetR, Cro is expressed from the trigger and represses both cI and Cro expression from the actuator. -2: Actuator in Cro state. expression of Xis and Int targets attachment sites on the trigger, excising *cro* and the terminator. +2: aTc represses TetR, mCherry is expressed from the trigger. Continued expression of Xis and Int prevents reintegration of excised DNA. **3C)** Response of the 434 two-counter circuit (n = 7) to an aTc induction time course with two stimulus pulses (red). aTc (100 ng/ml) was applied to liquid culture and washed out for 4 hour growth periods. A P_{tet}-mCherry strain (n = 4) is tested as a control (blue). Error bars represent SD. Spots of cultures on agar media (with or without aTc) are shown below at each step of the time course. Spot images consist of an mCherry color overlay on a grayscale brightfield image. **3D)** A 3-element Two-counter circuit that activates a memory element to register and record two exposures to aTc. **3E)** Response of 434 two-counter memory circuit and 434 two-counter circuit to an aTc induction time course with two stimulus pulses, followed by a final growth step in the absence of stimulus. aTc (100 ng/ml) was subsequently applied and washed out for growth periods of at least 4 hours. Error bars represent SD of three biological replicates. Spots of cultures on agar media containing X-gal (with or without aTc) are shown below at each step of the time course, full-color brightfield image.

and attR in the trigger, leaving an attB scar site (Figure 3B, -2 step). As a result of this excision, the trigger is primed for the second count, and a second aTc induction leads to the expression of the mCherry reporter. The continued presence of Xis prevents further re-integration (Abremski & Gottesman, 1982; Better et al., 1983) (Figure 3B, +2 step. -1 = before first pulse of aTc, +1 = during the first pulse of aTc, -2 = After first pulse/before the second pulse of aTc, +2 = during the second pulse of aTc.

The two-counter strain was tested using the P_{tet} promoter controlling the trigger and applying a time course of sequential pulses of aTc induction and quantifying the response of the strain via mCherry fluorescence (Figure 3C). The two-counter showed robust counting behavior, with two pulses required to produce mCherry fluorescence. There was virtually no mCherry signal during or after the first aTc pulse, even when the pulse was applied for up to 24 hours. At the second aTc application, a high percentage of cells displayed fluorescence (-aTc1, +aTc1, and -aTc2: 0.0% ± 0.0% SD; +aTc2: 95.9% ± 3.3% SD; n = 12). The excision of the two-counter trigger was confirmed by PCR amplification of the

trigger region at each step of the time course. Gel electrophoresis of the PCR product showed a reduction in trigger length consistent with excision of the region between attL and attR (full trigger length: 4.0 kb; excised trigger length: 3.3 kb) (Figure S3A). Furthermore, analysis by PCR confirmed that any mCherry-negative colonies had not undergone excision and still had the original trigger.

The two-counter also displays a high signal-to-noise ratio (Figure S3B) due to the stability of the phage 434 lysis–lysogeny switch in the cI state. The 434 cI protein used to maintain the memory-off state contains an ind- mutation, based on ind- mutants in phage λ (Gimble & Sauer, 1985). This mutation prevents RecA-mediated cleavage—the typical mechanism of induction in wild-type lambdoid phages during an SOS response (Sauer et al., 1982) —allowing it to stably maintain the cI state prior to the first application of aTc.

The bistable nature of the cI–Cro switch also produces a digital output with a clear demarcation in fluorescence between mCherry-negative and mCherry-positive cells (Figure S3B). Prior to the second application of aTc, all cells maintained a fluorescence below ~100 a.u. At the second application of aTc, a bimodal population was observed, in which there was a marked increase in mCherry fluorescence for cells that have completed trigger excision as a result of the removal of the first aTc stimulus. Because the population of cells that have experienced two pulses does not overlap with the population of cells experiencing fewer than two pulses, the readout of the count can be inferred at a single-cell level. In addition to the aTc-responsive two-counter based on the phage 434 OR operon, we also constructed a similar circuit based on the OR operon from phage lambda (Figures S3C-F). These results demonstrate our ability to construct falling-edge pulse counters based on lambdoid phage switches with high stability and signal-to-noise ratio.

A key motivation for constructing sensing circuits in bacteria is the potential to deploy them in inaccessible environments, such as the mammalian gut. For instance, bacteria containing a pulse

counter might be used to track the number of times they encounter a signal during transit through the intestine. After exiting the gut, the bacteria may be recovered to obtain a non-invasive readout of the registered count. While the two-counter described above produces distinct responses to two discrete pulses of a stimulus, its reporter readout requires active presence of that stimulus, as is the case with previous pulse counter circuits (Friedland et al., 2009). For a deployment-and-recovery scenario, a circuit must be able to record the count and produce a sustained readout which is interpretable even when the sensor is removed from the sensing environment.

We constructed a two-counter memory circuit (tc-P_{tet}-memory), which allows both sensing and recording of two discrete stimulus pulses (Figure 3D). The tc-P_{tet}-memory consists of a trigger, an actuator based on the phage 434 cI–Cro switch, and a memory switch based on the phage λ cI–Cro switch (Jonathan W Kotula et al., 2014; Naydich et al., 2019; Riglar et al., 2017). With the first aTc application, the circuit behaves identically to the 434 two-counter, producing Xis and Int from the actuator with the falling edge of the first pulse. On the second aTc application, the trigger produces λ cIDN (Naydich et al., 2019), which flips the λ memory switch to the Cro (memory-on) state so that it continuously produces a LacZ reporter. As with the previous two-counter, the two-counter memory circuit is designed to respond to the falling edge of the first stimulus pulse and the rising edge of the second.

The two-counter memory strain was subjected to a time course of aTc pulses, with a third aTc-free step after the second aTc pulse. The response at each step was quantified by plating cultures on X-gal indicator plates (+/- aTc, as appropriate) (Figure 3E). The strain showed nearly zero activation prior to the second pulse of aTc. The second aTc application resulted in a high percentage of cells in the on state, which persisted even after aTc was removed (-aTc1: 0.0% ± 0.0% SD; +aTc1: 3.2% ± 4.1% SD; -aTc2: 0.0% ± 0.0% SD; +aTc2: 88.0% ± 6.3% SD; -aTc3: 88.7% ± 4.1% SD; n = 3). The two-counter memory

circuit demonstrates our ability to incorporate complex counting and recording logic by combining multiple lambdoid phage switches in a synthetic context.

pH sensitive counting and containment mechanisms

The pulse counting systems developed for P_{tet} were reconstructed with P_{asr} controlling expression of *cro* and *gfp* (Figure 4A), yielding similar results. A plasmid with the P_{asr} promoter from acidTRP-09 controlling expression of a trigger element with GFP as the final reporter gene was transformed into an *E. coli* K12 strain containing the actuator, creating strain tc-P_{asr}AT09:GFP (Figure 4A). This strain was assayed by applying a time course of 5 hour growth steps alternating between pH 7 and pH 5. There was virtually no GFP signal before, during or after the first exposure to pH 5. Upon the second exposure to pH 5, almost 100% of cells displayed fluorescence (-pH5 1: 0.0% ± 0.0% SD, +pH5 1: 0.05% ± 0.03% SD, -pH5 2: 0.18% ± 0.04 SD; +pH5 2: 98.83 % ± 0.82% SD; n = 3) (Figure 4B).

Sustained exposure to pH 5 and pH 7 did not cause unwanted expression of GFP. tc-P_{asr}AT09:GFP was cultured overnight in pH 7 media, and then exposed to three five-hour growth steps in every possible permutation of pH 7 and pH 5 media (Figure S4A). Even when cultured in pH 5 media for 15 hours (approximately 30 generations), negligible GFP expression was observed. GFP expression was limited to only the culture that had been exposed to pH 5 conditions for two nonconsecutive growth steps (Figure S4B).

A pH sensitive two-count memory circuit using LacZ as a reporter was constructed (tc-P_{asr}AT09-memory) (Figure 4C). This strain was subjected to a time course with varying pH levels, after which memory response was quantified by plating on pH 7 X-gal indicator plates. Approximately 5% of cells showed switching before the second exposure to pH 5, compared to effectively 100% after the second exposure (Figure 4D).

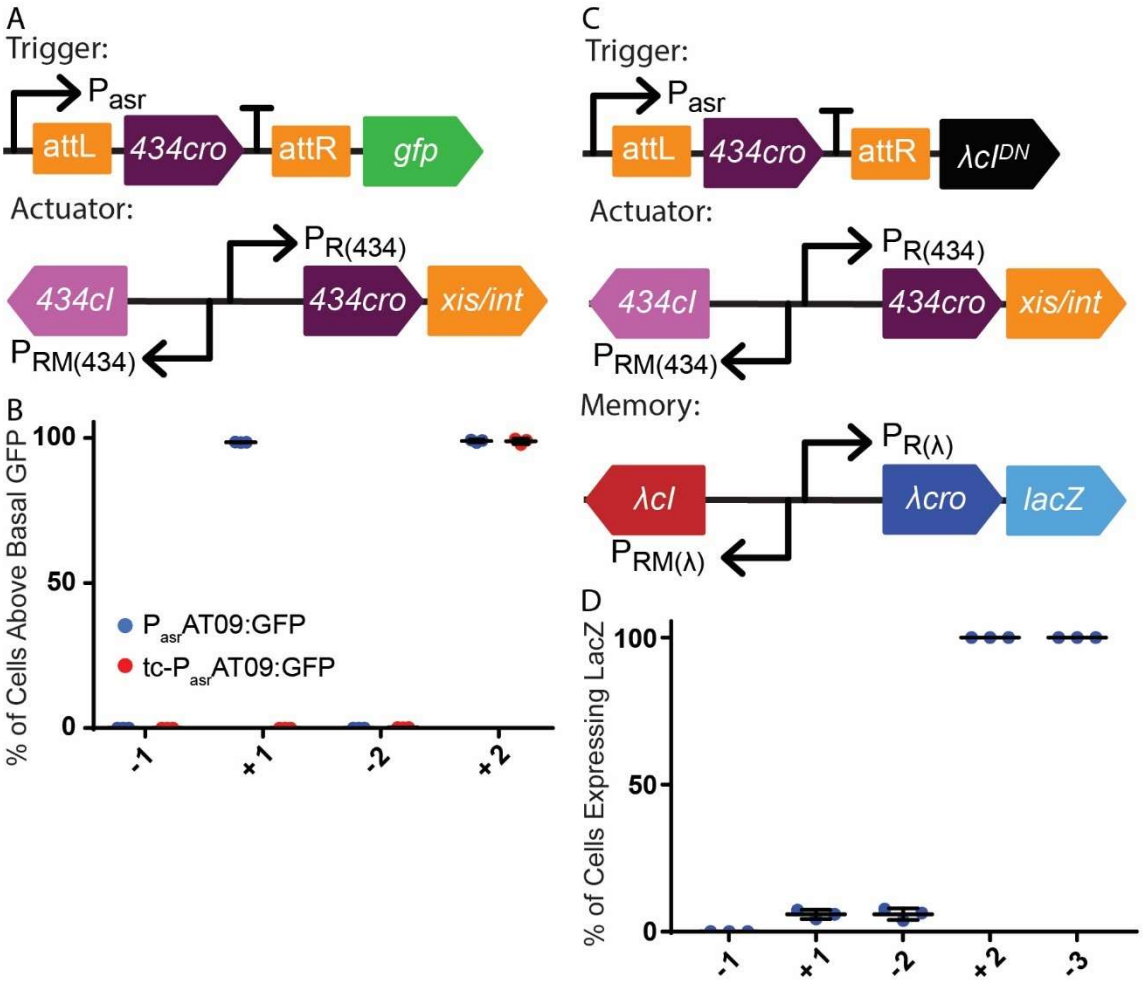

Figure 4. Analysis of the pH sensitive two counter (tc-P*asr*AT09:*gfp*) and two counter memory (tc-*Pasr*AT09-memory). Black arrows: promoter elements, attL/attR: 434 Xis/Int attachment sites, T: terminator, coloured arrows: genes. **4A)** Schematic of trigger and actuator of tc-P*asr*AT09:*gfp* strain. **4B)** percentage of cells expressing GFP above basal levels for P*asr*AT09:*gfp* (blue) and tc-P*asr*AT09:*gfp* (red). -1 = before first exposure to pH 5, +1 = during first exposure, -2 = after first exposure/ before second exposure, +2 = during second exposure. Data points from three biological replicates. Error bars represent the standard deviation. **4C)** Schematic of the trigger, actuator and memory cassette of tc-P*asr*AT09-memory strain. **4D)** The percentage of cells expressing lacZ for tc-P*asr*AT09-memory strain. -1 = before first exposure to pH 5, +1 = during first exposure, -2 = after first exposure/before second exposure, +2 = during second exposure, -3 = after the third exposure.

The pH-sensitive kill switch was combined with the pulse counter system to produce a kill switch that would only produce toxin upon a second exposure to low pH (tc-acidTRP-09). The toxin, *doc*, was inserted in place of a reporter on the trigger element, and the antitoxin, *phd*, was placed upstream of

P_{asr} (Figure 5A). After a single exposure to pH 5 and subsequent growth at pH 7, *434 cro* is excised, and the trigger element of the resulting circuit becomes identical to the kill switch circuit displayed in Figure 1A, aside from the presence of an attB scar site between P_{asr} and the toxin RBS (Figure 5B).

tc-acidTRP-09 displayed a survival ratio of 10^{-3} to 10^{-4} after undergoing a time course of pH 7 and pH 5 conditions. Length of exposure to the different pH levels was explored to identify the induction conditions resulting in the lowest survival ratio, and a pattern of 5 hours at pH 7 (-1), 10 hours at pH 5 (+1), 5 hours at pH 7 (-2) and finally 5 hours at pH 5 (+2) was settled on, all at 37 ^{o}C (data not shown for alternate conditions). The initial growth at pH 7 could also be an overnight growth at room temperature without affecting the resulting survival ratio (data not shown). After each growth step, serial dilutions of each culture were plated on both pH 7 and pH 5 plates, and a survival ratio was calculated. After the -1 and +1 steps, the log survival ratio was approximately 0, indicating excision had not taken place and *doc* was not being expressed. After the -2 step, a log survival ratio of approximately -4 was observed, indicating excision had occurred. After the +2 step, any surviving cells were plated, and the log survival ratio had returned to approximately 0. This result is expected, presumably because any cells remaining at the end of the +2 step had survived due to a mutation preventing tc-acidTRP-09 from functioning or because they had not performed trigger excision during the previous step—or a mixture of both (Figure 5C). Accordingly, the total cfu in the culture at the end of the +2 step was dramatically reduced (Figure 5C), indicating only a small proportion of the population was still viable. Surprisingly, PCR evidence showed that 100% (n=7) of escapees on pH 5 plates had excised, suggesting that toxin mutation was the primary reason for escape in this context.

tc-acidTRP-09 only displayed a decrease in survival ratio when exposed to pH 5 conditions on two separate, non-consecutive occasions. After an initial growth at pH 7, tc-acidTRP-09 was grown for at least 5 hours at either pH 7 or pH 5 for two consecutive growth steps in each possible permutation, represented by the 2nd and 3rd growth steps in Figure S4A. A serial dilution of the resultant cultures was

then plated on pH 7 and pH 5 plates, and a survival ratio was calculated for each set of conditions (Figure 5D). Sustained exposure to pH 5 and pH 7 failed to drop the survival ratio below 0, whereas a single exposure to pH 5 followed by a growth period at pH 7 dropped the survival ratio to approximately $10^{-3.5}$. This is expected, as growth on the pH 5 plate would be the second exposure to pH 5 conditions and would therefore induce expression of the toxin only in strains that had excised the *cro* element.

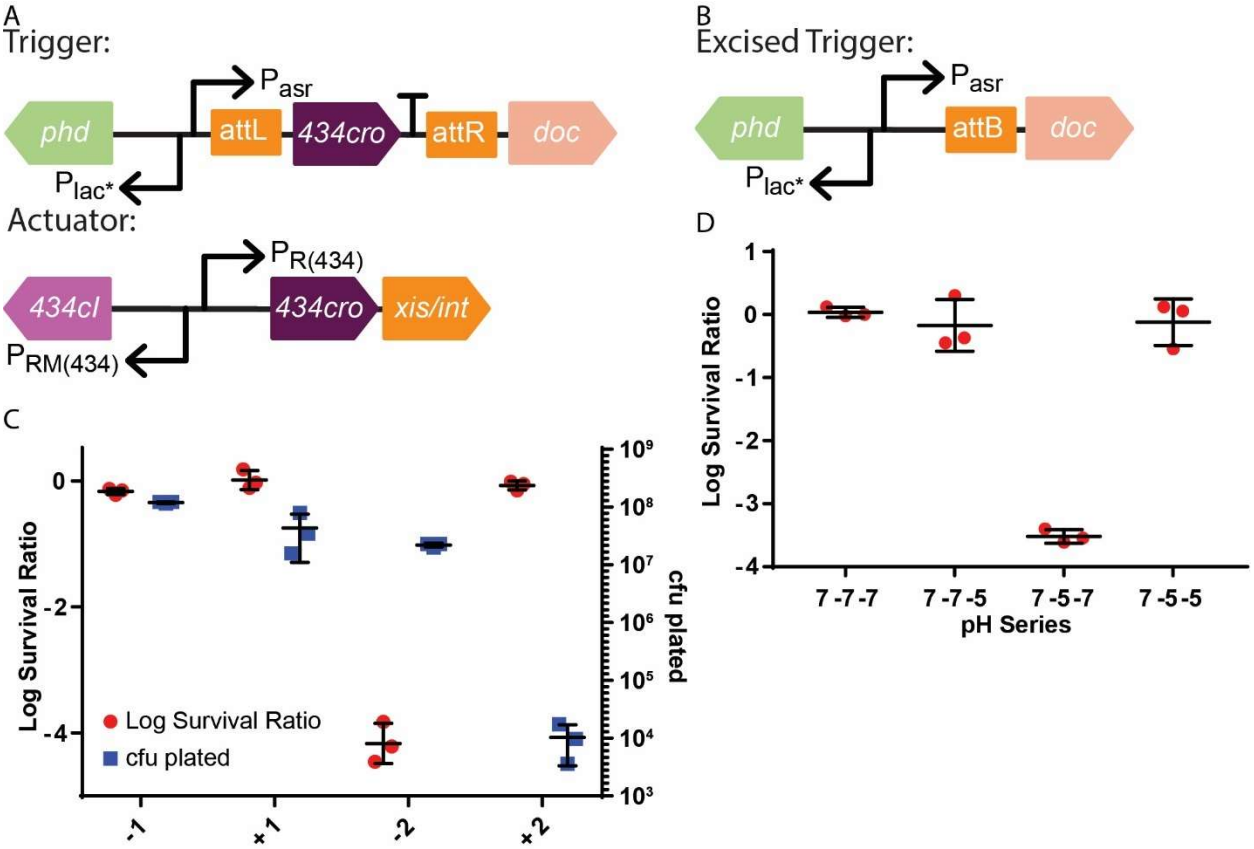

Figure 5. analysis of the two count pH sensitive kill switch (tc-acidTRP-09). Black arrows: promoter elements, attL/attR: 434 Xis/Int attachment sites, T: terminator, coloured arrows: genes. **5A)** Schematic of the trigger and actuator of tc-acidTRP-09. **5B)** Schematic of the trigger of tc-acidTRP-09 after excision of 434cro element. **5C)** Survival ratio (red) and total cfu on pH 7 plates (blue) of tc-acidTRP-09 throughout a time course experiment of varying pH conditions. -1 = before first exposure to pH 5, +1 = during first exposure, -2 = after first exposure and +2 = during second exposure. **5D)** Log survival ratio of tc-acidTRP-09 after exposure to 5 hours at 37 °C in three consecutive conditions. pH 7 followed by either pH 7 or pH 5 followed by either pH 7 or pH 5. After exposure to the final condition, cultures were plated on pH 7 and pH 5 plates and a survival ratio was calculated. Only the 7-5-7 culture plated on pH 5 had two, nonconsecutive exposures to pH 5 conditions.

tc-P_{asr}-memory was not functional in BalbC mouse model. Published values for mouse digestive pH begin at pH 3-4 in the stomach before rising to approximately pH 5-5.2 from the duodenum to the distal colon (Mcconnel et al 2008). Therefore a population that resides in the digestive tract should experience activation of the Pasr trigger element. tc-P_{asr}-memory was gavaged into mice and fecal samples were taken at various time points over the next 48 hours. However subsequent plating on M9 x-gal indicator plates yielded no blue colonies (data not shown), unlike the in vitro assay where 5% of colonies displayed lac Z activity (Figure 4D). Individual colonies from the 24 hour time point were then cultured in pH 5 media to ascertain whether the system was still functional. When put through the normal time course assay of pH 7 and pH 5 conditions, they behaved as expected, expressing lacZ in 100% of the population only after two exposures to pH 5.

The mouse digestive tract is not acidic enough to activate P_{asr}-. Two hypothesis existed for the lack of Pasr trigger activation in a mouse model. First, it is possible that the pH of the digestive tract in our BALB/c mice was not as low as the reported pH and therefore too high for P_{asr} activation. Alternatively, the mouse digestive tract contained an inhibiting factor that prevented expression from P_{asr}. To investigate these possibilities, we extracted and pooled cecal contents from 6 individual mice. These were resuspended in pH 5 media, which resulted in a final pH of 5.5. A time course alternating between pH 7 and pH 5 media containing resuspended cecal contents, both at pH 5.5 and readjusted to pH 5, was conducted. Excision and switching to LacZ expression was observed in the pH 5 media plus cecum fluid, but not in the pH 5.5 media with cecum fluid. This indicates that at least for our mouse model, their digestive pH is not consistently low enough to activate the P_{asr} trigger.

4.4 Discussion

We have used the *E. coli* pH sensitive promoter P_{asr} to construct toxin/anti-toxin based containment system called acidTRP. This containment system, which uses a kill switch activated at low-pH conditions, was shown to be evolutionarily stable for at least 100 generations and to result in a survival ratio of less than 10^{-6} upon exposure to pH 5. To achieve these attributes, we constructed and screened a rationally designed toxin/antitoxin library with varying levels of expression dictated by degenerate bases in the promoter RNAP -10 and ribosome binding sites of both the toxin and the antitoxin. This method was previously used to build the temperature sensitive kill switch cryodeath (Stirling et al., 2017), and the successful construction of acidTRP demonstrates the broader application of the method.

acidTRP was successfully combined with cryodeath into a single strain that responded to exposure to both low pH and low temperature. Respectively, each containment system conferred a survival ratio of 10^{-6} and 10^{-5}. The combined strain displayed a survival ratio below the limit of detection of 10^{-11}, indicating the two systems combined in a multiplicative manner. This is to be expected, as cryodeath uses an orthogonal type II toxin/antitoxin system, CcdB/CcdA, with a different cellular target (Madl et al., 2006; Wright et al., 2013), limiting interaction between the two kill switches and insuring that an individual cell must contain disabling mutations for both systems in order to survive. In addition, it has been shown that there is little cross talk between the regulation of P_{asr} and the cold shock response promoter P_{cspA} used in cryodeath (Hoynes-O'Connor et al., 2017).

For effective control of a bacterial population, a containment system's survival ratio must be low enough that the probability of an escape event over the time period of the bacterial population's intended use is negligible. In the example of the human digestive system, an individual expels *E. coli* at a rate of approximately 10^7 cfu per gram of excrement (Farnleitner et al., 2011) and produces approximately 300 g of excrement per day (Hosseini, 2000). Therefore, for a population of *E. coli*

residing in the human digestive tract over the course of one year, a survival ratio of less than 10^{-11} would result in fewer than 10 escape events. Whilst this does not take into account any potential fitness defect for the bacteria accrued in the process, the current standard for containment of transgenic *E. coli* host vectors is set by the NIH as no more than 1 in 10^8 surviving to pass on genetic code under specified conditions (NIH Guidelines, 2019, Apendix I-I-B). Whilst our containment system meets this standard, it is possible a more stringent standard is required for applications of *E. coli* outside of a laboratory.

In addition to the combined pH and temperature sensitive kill switches, we constructed pulse-sensing and pulse-counting circuits, based on a bi-stable, bacteriophage-derived switch with a built-in negative feedback mechanism. Uniquely for synthetic biological circuits, we have shown that our counters can respond to the falling edge of a stimulus pulse, a critical feature for circuits designed to produce unique responses to multiple discrete pulses of the same stimulus.

Our circuits demonstrated robust counting behaviour *in vitro*, exhibiting three distinct advantages over previous counters. First, our counters record only a single count for a single pulse, even when pulses were applied for up to 15 hours across multiple back dilutions (Figures S4B and 5D). Because the sustained intermediate state of the cI–Cro switch can be maintained indefinitely while the trigger is activated, we posit that there is no upper limit to the concentration and duration of an applied pulse that would cause the circuit to record more than a single count. Second, our counters have a tightly controlled off-state, ensuring no spontaneous switching, resulting in a high signal-to-noise ratio (Figures 3C, 3E, 4B, 4D, 5C and 5D). Third, the counters are digital reporters, which enables an interpretable readout of the count on a single-cell basis. This improves upon previous pulse counters, which produced reporting states that overlap between populations experiencing varying numbers of pulses, and thus were only interpretable when sampled as a population (Friedland et al., 2009). Single-cell readable counts are particularly important for applications in which there may be a limited number

of counter cells recovered from the environment, and therefore no ability to analyse population-level signals.

By combining two phage-derived bidirectional switches in a single circuit, we created a two-counter with memory, which continuously produces a reporter after a cell has experienced the falling edge of a first pulse followed by the rising edge of a second. Like the 434 and λ two-counters, this circuit showed efficient counting *in vitro (*Figures 3E and 4D).

The ability to reliably respond to distinct stimulus pulses enables a range of potential applications in tracking and responding to extracellular and intracellular signals. We have shown the counter can easily be adjusted to respond to a variety of sensors by altering the promoter controlling expression of *cro* on the trigger, such as by using the P_{asr} promoter. Synthetic biology has taken advantage of natural and designed parts to construct complex circuits, which incorporate increasingly sophisticated logic. Our demonstration of falling-edge pulse counters expands the repertoire of synthetic circuit functions and provides a scalable design for higher-order sensing, while our integration of kill switch and counter circuits demonstrates the use of such sensing to program specific cellular behaviours.

Acknowledgements

We owe a lot to the dedicated work of Loden Dundutsang, producing the many thousands of pH buffered plates this work required. We thank David Riglar for consultations on designing and implementing the pulse counting systems. This work was supported by Defense Advanced Research Projects Agency grant [HR0011-15-C-0094] and funds from the Wyss Institute for Biologically Inspired Engineering. F.S. acknowledges funding from NIH training grant [5T32GM007598]. A.N. acknowledges funding from National Science Foundation Graduate Research Fellowship [DGE1144152 and DGE1745303].

Author Contributions

Conceptualization, F.S., A.N., M.C, H.W. and J.W.; Investigation, F.S., A.N., J.B., R.B., M.C., H.W., E.R., S.O., S.G., A.C., Writing – Original Draft, F.S., A.N., Writing – Review & Editing, F.S., A.N., J.W., P.A.S., Funding Acquisition, P.A.S. and J.W.

4.5 Materials and Methods

EXPERIMENTAL MODEL AND SUBJECT DETAILS

***E. coli* K12 strain MG1655**

Used as the basis for all strains except for the tc-memory strains (both aTc and pH sensitive). Maintained using established protocols for *E. coli*. Strains that contained acidTRP were maintained in buffered pH 7 media, unless being assayed, and were prevented from growing above OD 1.5 to prevent acidification of the media. Strains that contained cryodeath candidates were maintained at a constant temperature of 37 ^{o}C unless being assayed.

***E. coli* K12 strain TB10**

A derivative of MG1655, with a large section of the lambda prophage genome inserted into the biotin operon. Used to recombine the acidTRP library and in strain construction for two counter circuits. cI has been mutated to the temperature sensitive variant, cI^{857}, allowing for temperature sensitive induction of the lambda red genes. Maintained using established protocols for E. coli, except all growth was kept at 30 ^{o}C.

***E. coli* strain NGF1**

E. coli NGF-1 is a strain originally isolated from the murine gut, which has proven to be an efficient and persistent colonizer and a reliable platform for the deployment of engineered circuits (Jonathan W Kotula et al., 2014).

METHOD DETAILS

Media and growth conditions:

Unless otherwise specified LB media with relevant antibiotic was used for all growth conditions. For the Buffered media, M9 minimal media supplemented with 1 mM MgSO4, 0.4% w/v glucose, 100 mM CaCl2 and 1 gram per liter of yeast extract (Difco) was used. For the pH 7 media the 5X M9 salts consisted of 15 mg/L CaCl2, 5 g/L NH4CL, 2.5 g/L NaCl, 30 g/L Na2HPO4, 15 g/L KH2PO4. For the pH 5 media the 5 x M9 salts consisted of 15 mg/L CaCl2, 5 g/L NH4CL, 2.5 g/L NaCl, 284 mg/L Na2HPO4, 43.4 g/L KH2PO4. Both were titrated to the desired pH using HCl or KOH. It should be noted that the pH 7 buffered media had a noticeably better buffering capacity. For the induction curve HCl or KOH where added to the pH 5 media (final pH 4.4 -5.8) and the pH 7 media (final pH 6). To make plates, 15 g/L of bacto agar was added. For all buffered media, distilled water was used, and it was observed that MillliQ ultrapure water resulted in minimal growth, presumably due to the lack of trace elements in the resultant media. Unless

otherwise stated, in all instances of antibiotic use the following concentrations were employed: Amp 100 µg/ml, Kan 50 µg/ml, Gent 10 µg/ml, Strep 100ug/ml, Cam 25 µg/ml. Unless otherwise specified, P_{tet} trigger strains were induced by adding 100 ng/ml aTc to liquid media or to agar plates. To quantify two-counter memory response on indicator plates, agar was supplemented with X-gal (60 µg/ml).

Acid sensitive reporter plasmid design

All primer sequences are given in Table S5. The P_{asr} regulatory region was ordered as reagents for a BioXP from SGI DNA, along with homology to the plasmid pUA66 (Table S1). pUA66 was linearised with the primers TS1 and TS2 and combined with the P_{asr} fragment using an NEB Gibson assembly kit according to the manufacturers guidelines, to create a plasmid with pH sensitive expression of GFP. The resultant plasmid was transformed into electrocompetent MG1655 cells.

Construction of the acidTRP library

The acidTRP library was ordered from SGI DNA as a custom order, with the 10 degenerate bases included (Table S2). On each end are 100 bp of homology to TrpC for homologous recombination into the genome. The resultant product was amplified using primers LS17 and LS18. After 35 cycles, the reaction was passed through a zymo clean and concentrator kit, and then run for 5 additional cycle with new PCR reagents. This step is to ensure that the final product contained two strands that matched completely, as mismatched strands may undergo mismatch repair in a strand independent manner, resulting in double stranded breaks as gap-repair polymerases meet.

P1 transduction

P1 lysates were constructed according to previously published methods (L. C. Thomason et al., 2007). Overnight cultures of donor strains were grown at relevant temperatures in LB, before being back diluted into 50 fold into 5 mL of LB supplemented with 0.2% w/v glucose, 20 mM MgCl2 and 5 mM CaCl2. After approximately 45 minutes of growth at the relevant temperature, between 10 and 100 ul of high titer (pfu > 100) lysate was added and the culture was grown until lysis had occurred. Lysates were then centrifuged at 4000 rpm for 5 minutes and the supernatant filtered using a 0.2 micron filter. Lysates were stored for up to 6 months at 4C.

Recombineering

Electrocompetent cells were prepared for transformation using previously published methods (L. Thomason et al., 2007). The acidTRP library was transformed into the recombinant strain TB10 that had been induced for 15 minutes in a 42 ^{o}C water bath with manual shaking every two minutes. Approximately 100 ng of DNA was combined with 50 ml of cells in a 0.1 cm cuvette, and electroporated using EC1 setting on a Biorad Micropulser. Cells were recovered in SOC medium for one hour before being spread on pH 7 buffered plates with kanamycin.

Initial acidTRP library screen

Approximately 200 colonies from the recombination of the library were struck out on both pH 7 and pH 5 plates. Any colony that showed a growth difference between the two conditions was struck to single colonies on both conditions, using cells from the pH 7 plate. Colonies that still displayed a growth difference between the two conditions were used to inoculate pH 7 liquid media and grown for 4 hours at 37 ^{o}C, and then used to make 25% glycerol stocks. All candidates were sequenced to determine the identity of each degenerate base (Table S3).

Constructing the toxin mutant control

The primer pairs LS17 and LS26 along with LS25 and LS18 were used to PCR the acidTRP library fragment from SGI DNA. The two resultant fragments were combined using an NEB Gibson assembly kit. The resultant product was transformed into TB10, and a single colony was picked to assay for lack of sensitivity to pH before being sequenced to confirm the frame shift mutation and identify the degenerate bases (Table S3).

pH sensitive survival assay

Glycerol stocks of acidTRP were used to inoculate pH 7 media and grown at 37 degrees until late log phase, approximately OD 1.0. One-hundred-fold serial dilutions of each culture were made in pH 7 media and plated onto both pH 7 and pH 5 buffered plates. Plates were then incubated for at least 48 hours at 37 ^{o}C. The cfu on each plate was counted and a ratio of pH 5 cfu to pH 7 cfu was calculated and termed the survival ratio.

Evolutionary stability of acidTRP strains

Glycerol stocks of acidTRP strains were used to inoculate 2 ml of pH 7 media, and then grown for approximately 24 hours at room temperature. These cultures were then sequentially passaged for another 20 growth steps by back diluting 100 fold into 2 ml of pH 7 media. OD was measured each day and used to calculate the number of generations that had passed. pH was measured after each growth step to ensure the pH of the media remained approximately at pH 7.

Competitive Growth Assay

Overnight cultures of MG1655 mixed with acidTRP-TM, 05, 09 and 10 were grown at room temperature to mid log phase. 100ul of the MG1655 culture was combined with each other culture, and then used to inoculate pH 7 media. Glycerol stocks were made of this initial inoculation. Cultures were passaged once every 24 hours diluting 1:1000 into pH 7 media. After 8 days, glycerol stocks were made of the final culture. Glycerol stocks were then defrosted, and a dilution series was plated onto both LB kanamycin and LB plates. The cfu ratio between the two types of plate was then used to determine the ratio of kill switch strains to MG1655.

Combined acidTRP and cryodeath survival assays

A P1 transduction was used to transfer each of the acidTRP-09 and acidTRP-TM cassettes into each of the cryodeath and cryodeath-TM strains, creating 4 new strains (acidTRP-TM/cryodeath-TM, acidTRP/cryodeath-TM, acidTRP-TM/cryodeath and acidTRP/cryodeath). For the first three, survival assays were conducted as before. For the acidTRP/cryodeath strain, glycerol stocks were used to inoculate 500 ml of pH 7 media and grown for ~ 10 hours at 37 ^{o}C until ~ OD 1. Cultures were centrifuged at 4000 rpm at 4 ^{o}C for 30 minutes in a 500 ml bucket centrifuge, resuspended in 20 ml of pH 7 media and transferred to 50 ml falcon tubes then spun down at 4000 rpm at 4 ^{o}C for 30 minutes. The supernatant was removed, and the pellet was resuspended in an additional 5 ml of pH 7 media (this represents the undiluted culture, 10^0). The undiluted culture was then diluted 10 fold followed by an additional 5 100 fold dilutions (10^1, 10^3, 10^5, 10^7, 10^9, 10^{11} dilutions). 225 *µl of e*ach dilution was then plated on two pH 7 and two pH 5 plates, and a plate of each pH was grown at both 37 ^{o}C and room temperature (~22 ^{o}C) for at least 5 days. 225 *µl* of the undiluted, 10^0 culture was also plated on 10 different pH 5 plates and grown at room temperature for the same time period. cfus were then counted in all conditions and a ratio of each compared to pH 7/37 ^{o}C was calculated as the survival assay for each condition.

Evolutionary stability of combined acidTRP and cryodeath strains

Glycerol stocks were used to innoculate 2 ml of pH 7 media and grown for approximately 4 hours at 37 ^{o}C, until mid-log phase. Cultures were then successively back diluted 100 fold into 2 ml of pH 7 media, and grown for 3 hours. 1-4 steps would be undertaken per day, and at the end of each day glycerol stocks would be made using 1:1 ratio of culture to 50% glycerol. In this way cultures were always kept at either 37 ^{o}C or -80 ^{o}C and were prevented from ever entering stationary phase and subsequently lowering the pH of the media.

Counter and memory strain construction

Actuator, memory and trigger elements were assembled using overlap extension PCR, Gibson assembly, or Golden Gate assembly. Actuator or memory elements were genomically integrated via λ Red recombineering into *E. coli* TB10. From *E. coli* TB10, actuator or memory elements were transduced by P1 transduction into *E. coli* K-12 MG1655. Trigger elements were either integrated into the genome via λ Red recombineering into a TB10 Δ*int/xis* mutant strain, and subsequently transduced by P1 transduction into the K-12 MG1655 strain containing the appropriate actuator or by assembling onto plasmids using a temperature-sensitive, Tn7 transposon insertion backbone derived from pGRG36. From these second group of plasmids, triggers were subsequently integrated into the K-12 MG1655 strain containing the appropriate actuator using the Tn7 transposon. For strain the tc-memory strains: The 434 actuator element was integrated into a TB10 strain via λ Red and subsequently P1 transduced into an *E. coli* NGF-1 strain that contains a λ memory element (Naydich et al., 2019). The trigger element was then integrated into this strain using the Tn7 transposon as described above.

Tn7 transposon integration

Trigger plasmids were electroporated into the appropriate actuator-bearing strain. Transformed cultures were grown in SOC media at 30°C for 90 minutes then resuspended in LB with ampicillin and grown at

30°C overnight to select transformants. Cultures were then back-diluted 1:100 into LB containing chloramphenicol and 0.1% arabinose to induce transposase genes while selecting for integration of triggers (which contained a chloramphenicol resistance cassette). After 6 hours at 30°C, cultures were back-diluted 1:100 into LB with chloramphenicol and grown at 42°C for at least 6 hours to cure the temperature-sensitive plasmid while selecting for integration. The back-dilution and curing were repeated a second time. Post-cure cultures were plated on agar containing chloramphenicol, and individual colonies were checked for integration at the attTn7 site by PCR and Sanger sequencing. Plasmid loss was confirmed by restreaming on agar plates containing ampicillin.

In vitro induction of aTc induced counter and memory strains

For induction time courses, strains were grown in liquid culture for periods of at least 4 hours at each step in the time course. aTc was added or removed from the media. Cultures were back diluted 1:100 between each step, with an extra wash prior to back dilution from +aTc to -aTc cultures. For some replicates, each step was as long as 24 hours. At the end of each step in the time course, 2 µl of each culture was spotted on an agar plate for imaging (with or without inducer and X-gal indicator, as appropriate). For strains with fluorescent reporters, cultures were analyzed by diluting 1:100 in PBS and measuring GFP and mCherry fluorescence using a BD LSR II Flow Cytometer. For strains with a LacZ reporter, cultures were plated on agar plates containing X-gal, and the strain response was assessed by counting blue (on) and white (off) colonies..

Imaging

Agar plates were imaged using a custom-built macroscope courtesy of the Kishony Lab and the Harvard Medical School Department of Systems Biology. All spot plates from induction time courses were imaged at 24 hours after plating. Consistent camera settings were used for each channel, and any contrast adjustments were applied consistently to all images in the series.

Counter strain trigger excision check by PCR

PCR was used to check for trigger excision as a result of counter induction. For The 434 two-counter with the trigger at the attTn7 locus, the primers used were ADN01 and AND02. For the λ two-counter with the trigger at the arabinose locus, the primers used were ADN03 and AND04. PCR products were run on an agarose gel with a 1 Kb Plus Ladder (ThermoFisher #10787018).

Construction of tc-P_{asr}AT09:GFP

Trigger element was constructed using a golden gate reaction of the products of three PCR reactions. Primer TC36 and TC37 using acidTRP-09 genomic DNA as a template, TC38 and TC39 using tc-P_{tet}:mCherry genomic DNA as a template, TC40 and TC47 using plasmid pFS01 (submitted to Addgene) as a template. The resultant plasmid was transformed into an *E. coli* K12 strain with the 434 actuator element integrated into the genome, resulting in strain tc-P_{asr}AT09:GFP.

Construction of tc-acidTRP09

The sequence in Table S4 was ordered as a gblock from Integrated DNA Technologies. This dsDNA was then transformed into TB10 using previously stated protocols, and the relevant region was transferred to into an *E. coli* K12 strain with the 434 actuator element integrated into the genome via P1 transduction, resulting in strain tc-acidTRP-09.

pH time course assay for tc-P$_{asr}$AT09:GFP two counter constructs.

tc-P$_{asr}$AT09:GFP and P$_{asr}$AT09:GFP glycerol stocks were used to inoculate pH 7 media and grown overnight at room temperature to late log phases. Cultures were then back diluted 1000 fold into either pH 7 or pH 5 media, and grown for 5 hours at 37 ^{o}C. cultures were again back diluted 1000 fold into pH 7 or pH 5 media and grown for 5 hours at 37 ^{o}C. A final 1 to 1000 back dilution into pH 7 or pH 5 media and 5 hour growth step at 37 ^{o}C was conducted. After each step a glycerol stock was made by combining cultures with 50% glycerol in a 1:1 ratio. Glycerol stocks were then defrosted and diluted 100 fold and assayed on a flow cytometer.

pH time course assay for tc-P$_{asr}$AT09-memory

tc-P$_{asr}$AT09-memory glycerol stocks were used to inoculate pH 7 media and grown overnight at room temperature to late log phase. four subsequent 5 hour growth steps at 37 ^{o}C were conducted, each one a successive 1000 fold dilution into alternating pH 5 then pH 7 media. Glycerol stocks were made after a growth step if it could not all be carried out on the same working day. After each growth step, a dilution series in the relevant media was made and plated onto pH 7 buffered plates supplemented with X-gal. the percentage of blue colonies was calculated and used to determine the percentage of cells expressing LacZ.

pH time course assay for tc-acidTRP-09

tc-acidTRP-09 glycerol stocks were used to inoculate pH 7 media and grown overnight at room temperature to late log phase. To assay the survival ratio after successive growth steps, overnight culture was back diluted 1000 fold into pH 5 media and grown for 5-10 hours at 37 ^{o}C. this was then back diluted 1000 fold into pH 7 media and grown for 5 hours at 37 ^{o}C, and then again into pH 5 media for a final 5 hour growth step at 37 ^{o}C. Glycerol stocks were made after a growth step if it could not all be carried out on the same working day. After each growth step, a dilution series in the relevant buffered media was made and plated on pH 7 and pH 5 plates as described for survival assays. To assay for survival ratio after varying sequences of growth conditions, the overnight culture was back diluted into both pH 7 or pH 5 media and grown for 5 hours at 37 ^{o}C for two more consecutive growth steps, resulting in four different sequences of conditions. Each of these four cultures was plated on both pH 7 and pH 5 plates, and a survival ratio was calculated for each.

Chapter 5 – Concluding remarks

5.1 Overview of work presented

With advances in synthetic biology, the ability to engineer bacteria, yeast, and other forms of microbial life has increased exponentially over the last few decades. The potential applications that arise from this new technology, while undoubtedly offering a plethora of positive benefits, can have both understood and unknown detrimental consequences when used in real world applications. This thesis aims to provide a summary of the field of biological containment for the application of genetically modified microorganisms, and outlines and demonstrates novel methods that can facilitate their application in uncontrolled environments. It covers techniques to:

- Create a genetic firewall between an engineered strain and wild-type organism, preventing the transfer and expression of transgenic material.
- Monitor the functional capacity of an engineered strain, initiating termination upon a loss of function mutation.
- Balance the expression of a toxin and antitoxin to produce an evolutionarily stable containment system.
- Construct a containment system that responds to an environmental signals rather than a small molecule inducer.
- Combine two environmentally sensitive containment systems into a single strain.
- Integrate a containment systems with a counting system designed to activate only upon a second exposure to a stimulus.

5.2 Summary of key achievements

Chapter 1 is a review of the field of biological containment. It outlines the history of containment, why biological containment is necessary and what new technologies have been produced or are being developmed that depend on effective containment strategies for safe application. A list of all systems developed for preventing the spread of plasmid born transgenic material is given. Methods for developing a genetic firewall between an engineered strain and other organisms are summarised, and recoding attempts to date are listed. Different techniques to control the physical translocation of transgenic microbes are discussed, summarizing the three main techniques of engineered auxotrophy, regulating an essential gene and regulating a toxic gene. 34 containment systems published over the last 4 decades are discussed, including an analysis of their modes of action and what scenarios they are applicable to. The implications of evolutionary stability, along with common mechanisms for transgenic organisms to escape their containment systems are considered. Finally, possible standards of practice for the field are posited.

Chapter 2 outlines a novel method, Stepwise Integration of Rolling Circle Amplified Segments (SIRCAS), to recode a bacterial genome using *Salmonella typhimurium* as a model organism. A recoded genome was generated removing all instances of the leucine encoding codons TTG and TTA, replacing them with CTG and CTA respectively. This would maintain a protein's primary sequence, but allow for the removal of the tRNA machinery associated with TTG/TTA. In addition, the genome was edited to remove all transposable elements, as well as pathogenicity islands and all instances of the recognition sequence for the type IIS restriction endonuclease LguI. Protocols for SIRCAS were designed, tested and implemented. 2-4 kb fragments of DNA were ordered from commercial sources and assembled into 10 – 20 kb fragments on yeast artificial chromosomes (YACs). These YACs were then amplified using rolling circle amplification, before being digested with LguI. The subsequent linear DNA was used to transform *S. typhimurium*, using the labda *red* recombinase machinery. Fragments were integrated on an iterative basis, with two different antibiotic selection markers being used to select for integration of the new

fragment while screening for removal of the old marker. Two separate genomic locations were recoded, before being combined into a single strain using CAGE (Lajoie, Rovner, et al., 2013). In total 200 kb of the 4.5 mb genome were recoded. Subsequent growth assays determined that the partially recoded strain had no significant growth defects.

Chapter 3 discusses two alternative methods for controlling the growth of a bacterial population. Both methods make use of the toxin, CcdB, expressed in response to an environmental signal. CcdB's natural antitoxin, CcdA, is expressed alongside the toxin, to remove any fitness defect that is caused by the leaky expression of the toxin in permissible conditions. Rationally designed libraries were constructed and screened to identify containment systems that had both a low survival ratio, and a high level of evolutionary stability. The first containment system, essentializer, uses a modified bacteriophage lambda P_R promoter to place the toxin under control of the transcription factors cI and Cro. If either transcription factor is expressed, the toxin is repressed. However if neither is present, toxin is expressed and cell death ensues. This system was shown to link cell survival to the presence of a second transgenic circuit, a memory element that exists in either the cI state or the Cro state (J W Kotula et al., 2014). The second containment system, cryodeath, uses the regulatory region from cold shock protein A to control expression of CcdB. At 37 ^{o}C, the toxin is repressed allowing cell survival, but at 22 ^{o}C toxin expression results in cell death. This system imparts a survival ratio of between 10^{-5} and 10^{-6} when a population transitions from 37 ^{o}C to 22 ^{o}C. Cryodeath was shown to maintain this survival ratio when used to contain a population to a mouse digestive system. Both the essentializer and cryodeath were shown to maintain their function over a period of 140 generations.

Chapter 4 shows the development of a third biological containment system, acid Termination of a Replicating Population (acidTRP). acidTRP uses the promoter from the acid shock response gene to regulate expression of the toxin Doc. Expression levels of Doc are balanced against expression levels of its antitoxin Phd using the same rationally designed library technique as in chapter 3. acidTRP was

shown to impart a survival ratio of 10^{-6} when a population transitioned from pH 7 to pH 5 growth conditions. acidTRP was combined with cryodeath into a single strain with a resultant combined survival ratio below the detection limit of 10^{-11}. In both instances it showed evolutionary stability over a period of 100 generations. Control of expression was increased by integrating it with a pulse counting system that responds to the falling edge of a stimulus pulse creating a two-count acidTRP (tc-acidTRP). tc-acidTRP displayed a survival ratio below 1 only when grown in pH 5 conditions for two, non-consecutive growth periods.

5.3 Future directions

Biological containment is an essential consideration when constructing transgenic strains for application in uncontrolled environments. Currently NIH guidelines recommend a survival ratio of less than 10^{-8} in order for a containment system to be considered effective (NIH FAQs, 2019). However, as postulated in the discussion of Chapter 4, this is likely not sufficient for many applications. A containment system should only be considered functional if its survival ratio is such that over the time period a transgenic microbe is to be deployed, the probability of an escape event occurring is negligible. For a transgenic population of microbes, if Y cells are expected to physically translocate away from their intended environment over the lifespan of the application, the survival ratio of a containment system must be orders of magnitude less than 10^{-Y}. This thesis does not look into what would be considered negligible, and future work on this matter would be informative.

The fitness of escapees also has a bearing on this equation. Although it is not investigated in this thesis, it is likely that the fitness of escapees from different forms of containment system will differ greatly. If a cells fitness is so reduced that it is quickly out competed and removed from an environment, it could be considered contained even if it escaped the initial containment system. Future work could assay the fitness of the escapees from the essentializer, cryodeath and acidTRP containment systems.

However, this premise should be used with caution. Examples such as transgenic mosquitos carrying lethal dominant genes hybridising with wild-type strains and surviving to sexual maturity serve as a cautionary tale (Evans et al., 2019).

While a survival ratio is an effective measurement of a containment system's capacity to control a bacterial population, the system's effectiveness is also dependent upon its capacity to maintain that survival ratio over a period of growth. This thesis focuses on mechanisms for maintaining evolutionary stability by reducing the fitness cost imparted by a containment system. An alternative approach would be to make a containment system impart a fitness benefit to its host strain. Recent work has explored the possibility of overlaying the ORF of a transgenic gene with that of an essential gene (Blazejewski et al., 2019). Using this technique to overlay the toxin of a containment system would make the presence of that toxin essential for the cell to function, therefore dramatically decreasing the potential avenues for escape.

5.4 Broader impact

Microbial biological containment is an enabling technology that allows the application of technologies that use transgenic microbes. Predominantly the different techniques outlined in this thesis can be directly or indirectly applied to the multitude of industrial, clinical and environmental applications that transgenic microbes have been proven or are predicted to have an impact in (talked abou tin section 1.1). As such , the broader implications of this thesis on the field are substantial.

In addition to the containment strategies, two techniques are explored that have a broader research application. First, a recoded organism can be a platform to address fundamental biological questions. New insight can be gained from studying cell response to codon replacements such as changes to global transcription/translation rates and the outcome of new biological function from incorporating non-natural amino acids into proteins. Novel methods for studying the transfer of genetic

information could be explored, or the effects of recoding on phage infection rates. Second, balancing the expression of two promoters was shown to be an effective method to create a stable containment system, but the technique of using small rationally designed libraries could be applied to other scenarios. Examples of systems that could all make use of the promoter variation technique explored in chapters 3 and 4 are system that depends on varying levels of two transcription factors, or optimal fluorescence from two reporters, or balancing the expression rates of a kinase/phosphatase pair.

This thesis seeks to be a framework that allows the development of transgenic microbes for applications in uncontrolled environments. By highlighting the importance of biological containment, outlining approaches to achieve the level of control necessary, and discussing the element of unknown that comes from applying these technologies, it encourages the discussion and development of innovative biological solutions to some of the multitude of problems the world currently faces.

References

Abremski, K., & Gottesman, S. (1982). Purification of the bacteriophage lambda xis gene product required for lambda excisive recombination. *Journal of Biological Chemistry*, *257*(16), 9658–9662. Retrieved from http://www.jbc.org.ezp-prod1.hul.harvard.edu/content/257/16/9658

Adrio, J. L., & Demain, A. L. (2010). Recombinant organisms for production of industrial products. *Bioengineered Bugs*, *1*(2), 116–131. https://doi.org/10.4161/bbug.1.2.10484

Agmon, N., Tang, Z., Yang, K., Sutter, B., Ikushima, S., Cai, Y., … Boeke, J. D. (2017). Low escape-rate genome safeguards with minimal molecular perturbation of Saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(8), E1470–E1479. https://doi.org/10.1073/pnas.1621250114

Ahrenholtz, I., Lorenz, M. G., & Wackernagel, W. (1994). A conditional suicide system in Escherichia coli based on the intracellular degradation of DNA. *Applied and Environmental Microbiology*, *60*(10), 3746–3751.

Angeles, L. (1976). *Heat-sensitive DNA-binding Activity of the cI Product of Bacteriophage Lambda*. *302*, 299–302.

Annaluru, N., Muller, H., Mitchell, L. A., Ramalingam, S., Stracquadanio, G., Richardson, S. M., … Chandrasegaran, S. (2014). Total Synthesis of a Functional Designer Eukaryotic Chromosome. *Science*, *344*(6179), 55–58. https://doi.org/10.1126/science.1249252

Arsène, F., Tomoyasu, T., & Bukau, B. (2000). The heat shock response of Escherichia coli. *International Journal of Food Microbiology*, *55*(1–3), 3–9. https://doi.org/10.1016/S0168-1605(00)00206-3

Atwood, B. Y. K. C. (1951). *(strain 15)*. 146–155.

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., … Mori, H. (2006). Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular Systems Biology*, *2*, 2006.0008. https://doi.org/10.1038/msb4100050

Babai, R., & Ron, E. Z. (1998). An Escherichia coli gene responsive to heavy metals. *FEMS Microbiology Letters*, *167*(2), 107–111. https://doi.org/10.1016/S0378-1097(98)00369-3

Baek, J. H., & Lee, S. Y. (2006). Novel gene members in the Pho regulon of Escherichia coli. *FEMS Microbiology Letters*, *264*(1), 104–109. https://doi.org/10.1111/j.1574-6968.2006.00440.x

Baeshen, N. A., Baeshen, M. N., Sheikh, A., Bora, R. S., Ahmed, M. M. M., Ramadan, H. A. I., … Redwan, E. M. (2014). Cell factories for insulin production. *Microbial Cell Factories*, *13*(1), 1–9. https://doi.org/10.1186/s12934-014-0141-0

Balan, A., & Schenberg, A. C. G. (2005). A conditional suicide system for Saccharomyces cerevisiae relying on the intracellular production of the Serratia marcescens nuclease. *Yeast*, *22*(3), 203–212. https://doi.org/10.1002/yea.1203

Baranyi, J., Metris, A., Carter, A. T., George, S. M., & Mulholland, F. (2014). Metabolic Shift of Escherichia coli under Salt Stress in the Presence of Glycine Betaine. *Applied and Environmental Microbiology*, *80*(15), 4745–4756. https://doi.org/10.1128/aem.00599-14

Barquist, L., Langridge, G. C., Turner, D. J., Phan, M.-D., Turner, A. K., Bateman, A., … Gardner, P. P. (2013). A comparison of dense transposon insertion libraries in the Salmonella serovars Typhi and

Typhimurium. *Nucleic Acids Research*, *41*(8), 4549–4564. https://doi.org/10.1093/nar/gkt148

Bej, A. K., Molin, S., Perlin, M., & Atlas, R. M. (1992). Maintenance and killing efficiency of conditional lethal constructs in Pseudomonas putida. *Journal of Industrial Microbiology*, *10*(2), 79–85. https://doi.org/10.1007/BF01583839

Bej, A. K., Perlin, M. H., & Atlas, R. M. (1988). Model suicide vector for containment of genetically engineered microorganisms. *Applied and Environmental Microbiology*, *54*(10), 2472–2477.

Berg, P., Baltimore, D., Brenner, S., Roblin, R. O., & Singer, M. F. (1975). Summary statement of the asilomar conference on recombinant DNA molecules. *Proceedings of the National Academy of Sciences of the United States of America*, *72*(6), 1981–1984. https://doi.org/10.1073/pnas.72.6.1981

Better, M., Wickner, S., Auerbach, J., & Echols, H. (1983). Role of the {Xis} protein of bacteriophage lambda in a specific reactive complex at the {attR} prophage attachment site. *Cell*, *32*(1), 161–168.

Bläsi, U., & Young, R. (1996). Two beginnings for a single purpose: the dual-start holins in the regulation of phage lysis. *Molecular Microbiology*, *21*(4), 675–682. https://doi.org/10.1046/j.1365-2958.1996.331395.x

Blazejewski, T., Ho, H., & Wang, H. H. (2019). Synthetic sequence entanglement augments stability and containment of genetic information in cells. *Science*, *365*(595–598), 1–5.

Boeke, J. D., Church, G., Hessel, A., Kelley, N. J., Arkin, A., Cai, Y., … Yang, L. (2016). The Genome Project-Write. *Science*. https://doi.org/10.1126/science.aaf6850

Braat, H., Rottiers, P., Hommes, D. W., Huyghebaert, N., Remaut, E., Remon, J. P., … Steidler, L. (2006). A Phase I Trial With Transgenic Bacteria Expressing Interleukin-10 in Crohn's Disease. *Clinical Gastroenterology and Hepatology*, *4*(6), 754–759. https://doi.org/10.1016/j.cgh.2006.03.028

Bremer, H. (1975). *Analysis of Enzyme Induction in Bacteria*. *152*, 243–254.

Burke, D., Carle, G., & Olson, M. (1987). Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science*, *236*(4803), 806–812. https://doi.org/10.1126/science.3033825

Cai, Y., Agmon, N., Choi, W. J., Ubide, A., Stracquadanio, G., Caravelli, K., … Boeke, J. D. (2015). Intrinsic biocontainment: multiplex genome safeguards combine transcriptional and recombinational control of essential yeast genes. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(6), 1803–1808. https://doi.org/10.1073/pnas.1424704112

Caliando, B. J., & Voigt, C. A. (2015). Targeted DNA degradation using a CRISPR device stably carried in the host genome. *Nature Communications*, *6*(May), 1–10. https://doi.org/10.1038/ncomms7989

Callura, J. M., Dwyer, D. J., Isaacs, F. J., Cantor, C. R., & Collins, J. J. (2010). Physiology Using Synthetic Riboregulators. *Proceedings of the National Academy of Sciences*, *107*(36), 15898–15903. https://doi.org/10.1073/pnas.1009747107

Caluwaerts, S., Vandenbroucke, K., Steidler, L., Neirynck, S., Vanhoenacker, P., Corveleyn, S., … Rottiers, P. (2010). AG013, a mouth rinse formulation of Lactococcus lactis secreting human Trefoil Factor 1, provides a safe and efficacious therapeutic tool for treating oral mucositis. *Oral Oncology*, *46*(7), 564–570. https://doi.org/10.1016/j.oraloncology.2010.04.008

Castro-Roa, D., Zenkin, N., Garcia-Pino, A., De Gieter, S., Van Nuland, N. A. J., & Loris, R. (2013). The Fic protein Doc uses an inverted substrate to phosphorylate and inactivate EF-Tu. *Nature Chemical Biology*, *9*(12), 811–817. https://doi.org/10.1038/nchembio.1364

Caulcott, C. A., & Rhodes, M. (1986). Temperature-induced synthesis of recombinant proteins. *Trends in Biotechnology*, Vol. 4, pp. 142–146. https://doi.org/10.1016/0167-7799(86)90164-2

Chan, C. T. Y., Lee, J. W., Cameron, D. E., Bashor, C. J., & Collins, J. J. (2015). "Deadman" and "Passcode" microbial kill switches for bacterial containment. *Nature Chemical Biology*, *12*(2), 82–86. https://doi.org/10.1038/nchembio.1979

Chari, R., & Church, G. M. (2017). Beyond editing to writing large genomes. *Nature Reviews Genetics*, *18*(12), 749–760. https://doi.org/10.1038/nrg.2017.59

Chatterjee, A., Lajoie, M. J., Xiao, H., Church, G. M., & Schultz, P. G. (2014). A bacterial strain with a unique quadruplet codon specifying non-native amino acids. *ChemBioChem*, *15*(12), 1782–1786. https://doi.org/10.1002/cbic.201402104

Chin, J. W. (2012). Reprogramming the Genetic Code. *Science*, *336*(6080), 428–429. https://doi.org/10.1126/science.1221761

Choi, Y., Sims, G. E., Murphy, S., Miller, J. R., & Chan, A. P. (2012). Predicting the Functional Effect of Amino Acid Substitutions and Indels. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0046688

Chou, C. H., Aristidou, a a, Meng, S. Y., Bennett, G. N., & San, K. Y. (1995). Characterization of a pH-inducible promoter system for high-level expression of recombinant proteins in Escherichia coli. *Biotechnology and Bioengineering*, *47*(2), 186–192. https://doi.org/10.1002/bit.260470210

Clardy, J., Fischbach, M. A., & Currie, C. R. (2009). The natural history of antibiotics. *Current Biology*, *19*(11), 1–8. https://doi.org/10.1016/j.cub.2009.04.001

Clark, R. L., Gordon, G. C., Bennett, N. R., Lyu, H., Root, T. W., & Pfleger, B. F. (2018). High-CO2 Requirement as a Mechanism for the Containment of Genetically Modified Cyanobacteria. *ACS Synthetic Biology*, *7*(2), 384–391. https://doi.org/10.1021/acssynbio.7b00377

Coley, W. B. (1893). The treatment of malignant tumors by repeated inoculations of erysipelas: With a report of ten original cases. *Clinical Orthopaedics and Related Research*.

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., … Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science*. https://doi.org/10.1126/science.1231143

Contreras, A., Molin, S., & Ramos, J. L. (1991). Conditional-suicide containment system for bacteria which mineralize aromatics. *Applied and Environmental Microbiology*, *57*(5), 1504–1508.

Cotter, P. A., Chepuri, V., Gennis, R. B., & Gunsalus, R. P. (1990). *Cytochrome o (cyoABCDE) and d (cydAB) Oxidase Gene Expression in Escherichia coli Is Regulated by Oxygen , pH , and the fnr Gene Product*. *172*(11), 6333–6338.

Csörgo, B., Fehér, T., Tímár, E., Blattner, F. R., & Pósfai, G. (2012). Low-mutation-rate, reduced-genome Escherichia coli: An improved host for faithful maintenance of engineered genetic constructs. *Microbial Cell Factories*, *11*, 1–13. https://doi.org/10.1186/1475-2859-11-11

Darling, A. E., Mau, B., & Perna, N. T. (2010). Progressivemauve: Multiple genome alignment with gene gain, loss and rearrangement. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0011147

Datsenko, K. a, & Wanner, B. L. (2000). One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(12), 6640–6645. https://doi.org/10.1073/pnas.120163297

de Lorenzo, V. (2009). Recombinant bacteria for environmental release: What went wrong and what we have learnt from it. *Clinical Microbiology and Infection*, *15*(SUPPL. 1), 63–65. https://doi.org/10.1111/j.1469-0691.2008.02683.x

Dean, F. B., Nelson, J. R., Giesler, T. L., & Lasken, R. S. (2001). Rapid amplification of plasmid and phage DNA using Phi29 DNA polymerase and multiply-primed rolling circle amplification. *Genome Research*. https://doi.org/10.1101/gr.180501

Dhall, P., Kumar, R., & Kumar, A. (2012). Biodegradation of sewage wastewater using autochthonous bacteria. *TheScientificWorldJournal*, *2012*. https://doi.org/10.1100/2012/861903

Diaz, E., Munthali, M., de Lorenzo, V., & Timmis, K. N. (1994). Universal barrier to lateral spread of specific genes among microorganisms. *Molecular Microbiology*, *13*(5), 855–861. https://doi.org/10.1111/j.1365-2958.1994.tb00477.x

Dirks, R. M., Bois, J. S., Schaeffer, J. M., Winfree, E., & Pierce, N. A. (2007). Thermodynamic Analysis of Interacting Nucleic Acid Strands. *SIAM Review*, *49*(1), 65–88. https://doi.org/10.1137/060651100

Diwo, C., & Budisa, N. (2019). Alternative biochemistries for alien life: Basic concepts and requirements for the design of a Robust biocontainment system in genetic isolation. *Genes*, *10*(1). https://doi.org/10.3390/genes10010017

Djordjevic, G. M., Sullivan, D. J. O., Walker, S. A., Conkling, M. A., & Carolina, N. (1997). *A Triggered-Suicide System Designed as a Defense against Bacteriophages †*. *179*(21), 6741–6748.

Englaender, J. A., Jones, J. A., Cress, B. F., Kuhlman, T. E., Linhardt, R. J., & Koffas, M. A. G. (2017). Effect of Genomic Integration Location on Heterologous Protein Expression and Metabolic Engineering in E. coli. *ACS Synthetic Biology*. https://doi.org/10.1021/acssynbio.6b00350

Etchegaray, J. P., & Inouye, M. (1999). Translational enhancement by an element downstream of the initiation codon in Escherichia coli. *Journal of Biological Chemistry*, *274*(15), 10079–10085. https://doi.org/10.1074/jbc.274.15.10079

Evans, B. R., Kotsakiozi, P., Costa-da-silva, A. L., Ioshino, R. S., Garziera, L., Pedrosa, M. C., … Powell, J. R. (2019). Transgenic Aedes aegypti Mosquitoes Transfer Genes into a Natural Population. *Scientific Reports*, 1–6. https://doi.org/10.1038/s41598-019-49660-6

Fang, L., Jiang, W., Bae, W., & Inouye, M. (1997). Promoter-independent cold-shock induction of cspA and its derepression at 37° C by mRNA stabilization. *Molecular Microbiology*, *23*(2), 355–364. https://doi.org/10.1046/j.1365-2958.1997.2351592.x

Farnleitner, A. H., Reischer, G. H., Burtscher, M. M., & Knetsch, S. (2011). *UKPMC Funders Group Author Manuscript indicators for human , livestock and wildlife faecal pollution in alpine mountainous water resources*. *109*(5), 1599–1608. https://doi.org/10.1111/j.1365-2672.2010.04788.x.Escherichia

Frahm, M., Felgner, S., Kocijancic, D., Rohde, M., Hensel, M., Curtiss, I. R., … Weiss, S. (2015). Efficiency of conditionally attenuated Salmonella enterica serovar typhimurium in bacterium-mediated tumor therapy. *MBio*. https://doi.org/10.1128/mBio.00254-15

Fredens, J., Wang, K., de la Torre, D., Funke, L. F. H., Robertson, W. E., Christova, Y., … Chin, J. W. (2019). Total synthesis of Escherichia coli with a recoded genome. *Nature*. https://doi.org/10.1038/s41586-019-1192-5

Friedland, A. E., Lu, T. K., Wang, X., Shi, D., Church, G., & Collins, J. J. (2009). Synthetic {Gene} {Networks} {That} {Count}. *Science*, *324*(5931), 1199–1202. https://doi.org/10.1126/science.1172005

Gallagher, R. R., Patel, J. R., Interiano, A. L., Rovner, A. J., & Isaacs, F. J. (2015). Multilayered genetic safeguards limit growth of microorganisms to defined environments. *Nucleic Acids Research*, *43*(3), 1945–1954. https://doi.org/10.1093/nar/gku1378

Geisow, J., & Evans, W. H. (1984). *pH in the Endosome exogenou*. *150*, 36–46.

Gibson, D. G., Benders, G. a, Axelrod, K. C., Zaveri, J., Algire, M. a, Moodie, M., … Hutchison, C. a. (2008). One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic Mycoplasma genitalium genome. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(51), 20404–20409. https://doi.org/10.1073/pnas.0811011106

Giege, R., Sissler, M., & Florentz, C. (1998). Universal rules and idiosyncratic features in tRNA identity. *Nucleic Acids Research*. https://doi.org/10.1093/nar/26.22.5017

Gimble, F. S., & Sauer, R. T. (1985). Mutations in bacteriophage lambda repressor that prevent {RecA}-mediated cleavage. *Journal of Bacteriology*, *162*(1), 147. Retrieved from http://jb.asm.org/content/162/1/147.abstract

Giuliodori, A. M., Di Pietro, F., Marzi, S., Masquida, B., Wagner, R., Romby, P., … Pon, C. L. (2010). The cspA mRNA Is a Thermosensor that Modulates Translation of the Cold-Shock Protein CspA. *Molecular Cell*, *37*(1), 21–33. https://doi.org/10.1016/j.molcel.2009.11.033

Guiney, D. G., & Yakobson, E. (1983). Location and nucleotide sequence of the transfer origin of the broad host range plasmid RK2. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.80.12.3595

Hankore, E. D., Zhang, L., Chen, Y., Liu, K., Niu, W., & Guo, J. (2019). Genetic Incorporation of Noncanonical Amino Acids Using Two Mutually Orthogonal Quadruplet Codons [Research-article]. *ACS Synthetic Biology*, *8*(5), 1168–1174. https://doi.org/10.1021/acssynbio.9b00051

Hirota, R., Abe, K., Katsuura, Z. I., Noguchi, R., Moribe, S., Motomura, K., … Kuroda, A. (2017). A novel biocontainment strategy makes bacterial growth and survival dependent on phosphite. *Scientific Reports*, *7*(March), 1–10. https://doi.org/10.1038/srep44748

Hoynes-O'Connor, A., Shopera, T., Hinman, K., Creamer, J. P., & Moon, T. S. (2017). Enabling complex genetic circuits to respond to extrinsic environmental signals. *Biotechnology and Bioengineering*, *9999*(xxx), 1–6. https://doi.org/10.1002/bit.26279

Huang, S., Lee, A. J., Tsoi, R., Wu, F., Zhang, Y., & Leong, K. W. (2016). *Coupling spatial segregation with synthetic circuits to control bacterial survival*. 1–13.

Hutchison, C. A., Chuang, R.-Y., Noskov, V. N., Assad-Garcia, N., Deerinck, T. J., Ellisman, M. H., … Venter, J. C. (2016). Design and synthesis of a minimal bacterial genome. *Science*, *351*(6280), aad6253–aad6253. https://doi.org/10.1126/science.aad6253

Isaacs, F. J., Carr, P. A., Wang, H. H., Lajoie, M. J., Sterling, B., Kraal, L., … Church, G. M. (2011). Precise

Manipulation of Chromosomes in Vivo Enables Genome-Wide Codon Replacement. *Science*, *333*(6040), 348 LP – 353. Retrieved from http://science.sciencemag.org/content/333/6040/348.abstract

Itaya, M., Tsuge, K., Koizumi, M., & Fujita, K. (2005). Combining two genomes in one cell: Stable cloning of the Synechocystis PCC6803 genome in the Bacillus subtilis 168 genome. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.0503868102

J, F. (1999). Intraluminal pH of the human gastrointestinal tract. *Danish Medical Journal*, *46*(3), 183–196.

Johnson, J. E., Lackner, L. L., Hale, C. a, Boer, P. a J. De, & Boer, a J. De. (2004). ZipA Is Required for Targeting of D MinC / DicB , but Not D MinC / MinD , Complexes to Septal Ring Assemblies in Escherichia coli ZipA Is Required for Targeting of D MinC / DicB , but Not MinC / MinD , Complexes to Septal Ring Assemblies in Escherichia co. *Society*, *186*(8), 2418–2429. https://doi.org/10.1128/JB.186.8.2418

Jones, P. G., Krah, R., Tafuri, S. R., & Wolffe, A. P. (1992). DNA gyrase, CS7.4, and the cold shock response in Escherichia coli. *J Bacteriol*, *174*(18), 5798–5802. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1325964

Karlinsey, J. E. (2007). λ-Red Genetic Engineering in Salmonella enterica serovar Typhimurium. *Methods in Enzymology*. https://doi.org/10.1016/S0076-6879(06)21016-4

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., … Drummond, A. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*. https://doi.org/10.1093/bioinformatics/bts199

Kelsic, E. D., Chung, H., Cohen, N., Park, J., Wang, H. H., & Kishony, R. (2016). RNA Structural Determinants of Optimal Codons Revealed by MAGE-Seq. *Cell Systems*. https://doi.org/10.1016/j.cels.2016.11.004

Khetrapal, V., Mehershahi, K., Rafee, S., Chen, S., Lim, C. L., & Chen, S. L. (2015). A set of powerful negative selection systems for unmodified Enterobacteriaceae. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkv248

Kipper, K., Lundius, E. G., Ćurić, V., Nikić, I., Wiessler, M., Lemke, E. A., & Elf, J. (2017). Application of Noncanonical Amino Acids for Protein Labeling in a Genomically Recoded Escherichia coli. *ACS Synthetic Biology*. https://doi.org/10.1021/acssynbio.6b00138

Knudsen, S. M., & Karlstrom, O. H. (1991). Development of efficient suicide mechanisms for biological containment of bacteria. *Applied and Environmental Microbiology*, *57*(1), 85–92.

Knudsen, S., Saadbye, P., Hansen, L. H., Collier, A., Jacobsen, B. L., Schlundt, J., & Karlstrom, O. H. (1995). Development and testing of improved suicide functions for biological containment of bacteria. *Applied and Environmental Microbiology*, *61*(3), 985–991.

Knuth, K., Niesalla, H., Hueck, C. J., & Fuchs, T. M. (2004). Large-scale identification of essential Salmonella genes by trapping lethal insertions. *Molecular Microbiology*, *51*(6), 1729–1744. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15009898

Koch, A. L. (1974). The pertinence of the periodic selection phenomenon to prokaryote evolution. *Genetics*, *77*(1), 127–142.

Kong, W., Brovold, M., Koeneman, B. A., Clark-Curtiss, J., & Curtiss, R. (2012). Turning self-destructing Salmonella into a universal DNA vaccine delivery platform. *Proceedings of the National Academy of Sciences*, *109*(47), 19414–19419. https://doi.org/10.1073/pnas.1217554109

Kong, Wei, Wanda, S., Zhang, X., Bollen, W., Tinge, S. A., Roland, K. L., & Curtiss, R. (2008). *Regulated programmed lysis of recombinant Salmonella in host tissues to release protective antigens and confer biological containment*. *105*(27), 9361–9366.

Kosuri, S., & Church, G. M. (2014). Large-scale de novo DNA synthesis: Technologies and applications. *Nature Methods*. https://doi.org/10.1038/nmeth.2918

Kotula, J W, Kerns, S. J., Shaket, L. A., Siraj, L., Collins, J. J., Way, J. C., & Silver, P. A. (2014). Programmable bacteria detect and record an environmental signal in the mammalian gut. *Proceedings of the National Academy of Sciences*, *111*(13), 4838–4843. https://doi.org/10.1073/pnas.1321321111

Kotula, Jonathan W, Kerns, S. J., Shaket, L. A., Siraj, L., Collins, J. J., Way, J. C., & Silver, P. A. (2014). Programmable bacteria detect and record an environmental signal in the mammalian gut. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(13), 4838–4843. https://doi.org/10.1073/pnas.1321321111

Kouprina, N., & Larionov, V. (2008). Selective isolation of genomic loci from complex genomes by transformation-associated recombination cloning in the yeast Saccharomyces cerevisiae. *Nature Protocols*. https://doi.org/10.1038/nprot.2008.5

Kristoffersen, P., Jensen, G. B., Gerdes, K., & Piškur, J. (2000). Bacterial toxin-antitoxin gene system as containment control in yeast cells. *Applied and Environmental Microbiology*, *66*(12), 5524–5526. https://doi.org/10.1128/AEM.66.12.5524-5526.2000

Kuhlman, T. E., & Cox, E. C. (2010). Site-specific chromosomal integration of large synthetic constructs. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkp1193

Kuo, J., Stirling, F., Lau, Y. H., Shulgina, Y., Way, J. C., & Silver, P. A. (2018). Synthetic genome recoding: new genetic codes for new features. *Current Genetics*, *64*(2). https://doi.org/10.1007/s00294-017-0754-z

Lajoie, M. J., Kosuri, S., Mosberg, J. A., Gregg, C. J., Zhang, D., & Church, G. M. (2013). Probing the limits of genetic recoding in essential genes. *Science*. https://doi.org/10.1126/science.1241460

Lajoie, M. J., Rovner, A. J., Goodman, D. B., Aerni, H.-R., Haimovich, A. D., Kuznetsov, G., … Isaacs, F. J. (2013). Genomically Recoded Organisms Expand Biological Functions. *Science*, *342*(6156), 357–360. https://doi.org/10.1126/science.1241459

Laurel A Lagenaur, Brigitte E Sanders-Beer, Beda Brichacek, Ranajit Pal, X., & Liu, Yang Liu, Rosa Yu, David Venzon, Peter P Lee, and D. H. H. (2011). Prevention of vaginal SHIV transmission in macaques by a live recombinant Lactobacillus. *Mucosal Immunol*, *4*(6), 648–657. https://doi.org/10.1038/nature22814.Trans-kingdom

Lee, H., Popodi, E., Tang, H., & Foster, P. L. (2012). Rate and molecular spectrum of spontaneous mutations in the bacterium Escherichia coli as determined by whole-genome sequencing. *Proceedings of the National Academy of Sciences*, *109*(41), E2774–E2783. https://doi.org/10.1073/pnas.1210309109

Lee, J. W., Chan, C. T. Y., Slomovic, S., & Collins, J. J. (2018). Next-generation biocontainment systems for engineered organisms. *Nature Chemical Biology*, *14*(6), 530–537. https://doi.org/10.1038/s41589-018-0056-x

Lee, S. J., Xie, A., Jiang, W., Etchegaray, J. P., Jones, P. G., & Inouye, M. (1994). Family of the major cold-shock protein, CspA (CS7.4), of Escherichia coli, whose members show a high sequence similarity with the eukaryotic Y-box binding proteins. *Molecular Microbiology*, *11*(5), 833–839. https://doi.org/10.1111/j.1365-2958.1994.tb00361.x

Li, Q., & Wu, Y. J. (2009). A fluorescent, genetically engineered microorganism that degrades organophosphates and commits suicide when required. *Applied Microbiology and Biotechnology*, *82*(4), 749–756. https://doi.org/10.1007/s00253-009-1857-3

Liu, M. A., Stent, T. L., Hong, Y., & Reeves, P. R. (2015). Inefficient translocation of a truncated O unit by a Salmonella Wzx affects both O-antigen production and cell growth. *FEMS Microbiology Letters*, *362*(9), fnv053–fnv053. https://doi.org/10.1093/femsle/fnv053

Liu, M., Zhang, Y., Inouye, M., & Woychik, N. A. (2008). Bacterial addiction module toxin Doc inhibits translation elongation through its association with the 30S ribosomal subunit. *Proceedings of the National Academy of Sciences*, *105*(15), 5885–5890. https://doi.org/10.1073/pnas.0711949105

Lopez, G., & Anderson, J. C. (2015). Synthetic Auxotrophs with Ligand-Dependent Essential Genes for a BL21(DE3) Biosafety Strain. *ACS Synthetic Biology*, *4*(12), 1279–1286. https://doi.org/10.1021/acssynbio.5b00085

Łukasiewicz, K., & Fol, M. (2018). Microorganisms in the Treatment of Cancer: Advantages and Limitations. *Journal of Immunology Research*, *2018*. https://doi.org/10.1155/2018/2397808

Ma, Natalie J, Moonan, D. W., & Isaacs, F. J. (2014). Precise manipulation of bacterial chromosomes by conjugative assembly genome engineering. *Nature Protocols*, *9*(10), 2285–2300. https://doi.org/10.1038/nprot.2014.081

Ma, Natalie Jing, & Isaacs, F. J. (2016). Genomic Recoding Broadly Obstructs the Propagation of Horizontally Transferred Genetic Elements. *Cell Systems*. https://doi.org/10.1016/j.cels.2016.06.009

Maddamsetti, R., Lenski, R. E., & Barrick, J. E. (2015). *Adaptation, Clonal Interference, and Frequency-Dependent Interactions in a Long-Term Evolution Experiment with Escherichia coli*. *200*(June), 619–631. https://doi.org/10.1534/genetics.115.176677

Madl, T., Van Melderen, L., Mine, N., Respondek, M., Oberer, M., Keller, W., … Zangger, K. (2006). Structural Basis for Nucleic Acid and Toxin Recognition of the Bacterial Antitoxin CcdA. *Journal of Molecular Biology*, *364*(2), 170–185. https://doi.org/10.1016/j.jmb.2006.08.082

Malan, T. P., & McClure, W. R. (1984). Dual promoter control of the escherichia coli lactose operon. *Cell*, *39*(1), 173–180. https://doi.org/10.1016/0092-8674(84)90203-4

Malyshev, D. A., Dhami, K., Lavergne, T., Chen, T., Dai, N., Foster, J. M., … Romesberg, F. E. (2014). A semi-synthetic organism with an expanded genetic alphabet. *Nature*, *509*(7500), 385–388. https://doi.org/10.1038/nature13314

Mandell, D. J., Lajoie, M. J., Mee, M. T., Takeuchi, R., Kuznetsov, G., Norville, J. E., … Church, G. M. (2015). Biocontainment of genetically modified organisms by synthetic protein design. *Nature*,

518(7537), 55–60.

Mapelli, F., Scoma, A., Michoud, G., Aulenta, F., Boon, N., Borin, S., … Daffonchio, D. (2017). Biotechnologies for Marine Oil Spill Cleanup: Indissoluble Ties with Microorganisms. *Trends in Biotechnology*, *35*(9), 860–870. https://doi.org/10.1016/j.tibtech.2017.04.003

Marlière, P., Patrouix, J., Döring, V., Herdewijn, P., Tricot, S., Cruveiller, S., … Mutzel, R. (2011). Chemical evolution of a bacterium's genome. *Angewandte Chemie - International Edition*, *50*(31), 7109–7114. https://doi.org/10.1002/anie.201100535

Matic, I., Rayssiguier, C., & Radman, M. (1995). Interspecies gene exchange in bacteria: The role of SOS and mismatch repair systems in evolution of species. *Cell*. https://doi.org/10.1016/0092-8674(95)90501-4

McClelland, M. (2000). Comparison of the Escherichia coli K-12 genome with sampled genomes of a Klebsiella pneumoniae and three Salmonella enterica serovars, Typhimurium, Typhi and Paratyphi. *Nucleic Acids Research*, *28*(24), 4974–4986. https://doi.org/10.1093/nar/28.24.4974

McClelland, Michael, Sanderson, K. E., Spieth, J., Clifton, S. W., Latreille, P., Courtney, L., … Wilson, R. K. (2001). Complete genome sequence of Salmonella enterica serovar Typhimurium LT2. *Nature*. https://doi.org/10.1038/35101614

McClure, N. C., & Venables, W. A. (1986). Adaptation of pseudomonas putida mt-2 to growth on aromatic amines. *Journal of General Microbiology*.

McConnell, E. L., Basit, A. W., & Murdan, S. (2007). Measurements of rat and mouse gastrointestinal pH, fluid and lymphoid tissue, and implications for in-vivo experiments. *Journal of Pharmacy and Pharmacology*, *60*(1), 63–70. https://doi.org/10.1211/jpp.60.1.0008

McKinley, J. E., & Magnuson, R. D. (2005). Characterization of the Phd repressor-antitoxin boundary. *Journal of Bacteriology*, *187*(2), 765–770. https://doi.org/10.1128/JB.187.2.765-770.2005

Menezes, A. A., Cumbers, J., Hogan, J. A., & Arkin, A. P. (2014). Towards synthetic biological approaches to resource utilization on space missions. *Journal of The Royal Society Interface*, *12*(102), 20140715–20140715. https://doi.org/10.1098/rsif.2014.0715

Mitchell, L. A., Wang, A., Stracquadanio, G., Kuang, Z., Wang, X., Yang, K., … Boeke, J. D. (2017). Synthesis, debugging, and effects of synthetic chromosome consolidation: synVI and beyond. *Science*, *355*(6329). https://doi.org/10.1126/science.aaf4831

Mitta, M., Fang, L., & Inouye, M. (1997). Deletion analysis of cspA of Escherichia coli: requirement of the AT-rich UP element for cspA transcription and the downstream box in the coding region for its cold shock induction. *Molecular Microbiology*, *26*(2), 321–335. https://doi.org/10.1046/j.1365-2958.1997.5771943.x

Moe-Behrens, G. H. G., Davis, R., & Haynes, K. A. (2013). Preparing synthetic biology for the world. *Frontiers in Microbiology*, *4*(JAN), 1–10. https://doi.org/10.3389/fmicb.2013.00005

Molin, S., Boe, L., Jensen, L. B., Kristensen, C. S., Givskov, M., Ramos, J. L., & Bej, A. K. (1993). Suicidal genetic elements and their use in biological containment of bacteria. *Annu Rev Microbiol*, *47*, 139–166. https://doi.org/10.1146/annurev.mi.47.100193.001035

Molin, S., Klemm, P., Poulsen, L., Biehl, H., Gerdes, K., & Andersson, P. (1987). Conditional suicide system

for containment of bacteria and plasmids. *Nature Biotechnology*, *5*(December), 1315–1318.

Montague, M., McArthur, G. H. th, Cockell, C. S., Held, J., Marshall, W., Sherman, L. A., … Cumbers, J. (2012). The role of synthetic biology for in situ resource utilization (ISRU). *Astrobiology*, *12*(12), 1135–1142. https://doi.org/10.1089/ast.2012.0829

Motieju, D. (2009). *Molecular Characterization of the Acid-Inducible asr Gene of Escherichia coli and Its Role in Acid Stress Response*. *185*(8), 2475–2484. https://doi.org/10.1128/JB.185.8.2475

Mulligan, M. E., Hawley, D. K., Entriken, R., & McClure, W. R. (1984). *Escherichia coli* promoter sequences predict in vitro RNA polymerase selectivity. *Nucleic Acids Research*, *12*(1), 789–800. https://doi.org/PMC321093

Murakami, K. S., Masuda, S., Campbell, E. a, Muzzin, O., & Darst, S. a. (2002). Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. *Science (New York, N.Y.)*, *296*(5571), 1285–1290. https://doi.org/10.1126/science.1069595

Napolitano, M. G., Landon, M., Gregg, C. J., Lajoie, M. J., Govindarajan, L., Mosberg, J. A., … Church, G. M. (2016). Emergent rules for codon choice elucidated by editing rare arginine codons in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.1605856113

Nathan, L. P., Glenn, D. L., & Johan, V. (2016). Synchronous long-term oscillations in a synthetic gene circuit. *Nature*, *538*(7626), 1–4. https://doi.org/10.1038/nature19841

Naydich, A. D., Nangle, S. N., Bues, J. J., Trivedi, D., Nissar, N., Inniss, M. C., … Riglar, D. T. (2019). Synthetic {Gene} {Circuits} {Enable} {Systems}-{Level} {Biosensor} {Trigger} {Discovery} at the {Host}-{Microbe} {Interface}. *MSystems*, *4*(4), e00125--19. https://doi.org/10.1128/mSystems.00125-19

NIH FAQs. (2013). FAQs NIH Guidelines for Research Involving Recombinant or Synthetic Nucleic Acid Molecules. *NIH Guidelines*, *2*(April), 1–7.

NIH FAQs. (2019). FAQs NIH Guidelines for Research Involving Recombinant or Synthetic Nucleic Acid Molecules. *NIH Guidelines*, *2*(April), 1–7.

Noman, N., Inniss, M., Iba, H., & Way, J. C. (2016). Pulse {Detecting} {Genetic} {Circuit} – {A} {New} {Design} {Approach}. *PLOS ONE*, *11*(12), e0167162. https://doi.org/10.1371/journal.pone.0167162

NOVICK, A., & SZILARD, L. (1950). Experiments with the Chemostat on spontaneous mutations of bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, *36*(12), 708–719. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1063276&tool=pmcentrez&rendertype=abstract

Ogasawara, H., Hasegawa, A., Kanda, E., Miki, T., Yamamoto, K., & Ishihama, A. (2007). Genomic SELEX search for target promoters under the control of the PhoQP-RstBA signal relay cascade. *Journal of Bacteriology*, *189*(13), 4791–4799. https://doi.org/10.1128/JB.00319-07

Ostrov, N., Landon, M., Guell, M., Kuznetsov, G., Teramoto, J., Cervantes, N., … Church, G. M. (2016). Design, synthesis, and testing toward a 57-codon genome. *Science*, *353*(6301), 819–822. https://doi.org/10.1126/science.aaf3639

Paridah, M. ., Moradbak, A., Mohamed, A. ., Owolabi, F. abdulwahab taiwo, Asniza, M., & Abdul Khalid, S. H. . (2016). Pesticides - Recent Trends in Pesticide Residue Assay. *Intech*, *i*(tourism), 13. https://doi.org/http://dx.doi.org/10.5772/57353

Pieper, D. H., & Reineke, W. (2000). Engineering bacteria for bioremediation. *Current Opinion in Biotechnology*, *11*, 262–270.

Piraner, D. I., Abedi, M. H., Moser, B. A., Lee-Gosselin, A., & Shapiro, M. G. (2016). Tunable thermal bioswitches for in vivo control of microbial therapeutics. *Nat Chem Biol*, *advance on*(November), 1–8. https://doi.org/10.1038/nchembio.2233

Poteete, A. R., & Fenton, A. C. (2000). Genetic requirements of phage λ red-mediated gene replacement in Escherichia coli K-12. *Journal of Bacteriology*, *182*(8), 2336–2340. https://doi.org/10.1128/JB.182.8.2336-2340.2000

Ptashne, M., Jeffrey, A., Johnson, A. D., Maurer, R., Meyer, B. J., Pabo, C. O., … Sauer, R. T. (1980). How the lambda repressor and cro work. *Cell*, *19*(1), 1–11. https://doi.org/10.1016/0092-8674(80)90383-9

Razzaq, A., Shamsi, S., Ali, A., Ali, Q., Sajjad, M., Malik, A., & Ashraf, M. (2019). Microbial proteases applications. *Frontiers in Bioengineering and Biotechnology*, *7*(JUN), 1–20. https://doi.org/10.3389/fbioe.2019.00110

Recorbet, G., Robert, C., Givaudan, A., Kudla, B., Normand, P., & Faurie, G. (1993). Conditional suicide system of Escherichia coli released into soil that uses the Bacillus subtilis sacB gene. *Applied and Environmental Microbiology*, *59*(5), 1361–1366.

Reeves, P. R., Cunneen, M. M., Liu, B., & Wang, L. (2013). Genetics and Evolution of the Salmonella Galactose-Initiated Set of O Antigens. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0069306

Richardson, S. M., Mitchell, L. A., Stracquadanio, G., Yang, K., Dymond, J. S., DiCarlo, J. E., … Bader, J. S. (2017). Design of a synthetic yeast genome. *Science*, *355*(6329), 1040–1044.

Riglar, D. T., Giessen, T. W., Baym, M., Kerns, S. J., Niederhuber, M. J., Bronson, R. T., … Silver, P. A. (2017). Engineered bacteria can function in the mammalian gut long-term as live diagnostics of inflammation. *Nature Biotechnology*, *35*(7), nbt.3879. https://doi.org/10.1038/nbt.3879

Riglar, D. T., & Silver, P. A. (2018). Engineering bacteria for diagnostic and therapeutic applications. *Nature Reviews Microbiology*, *16*(4), 214–225. https://doi.org/10.1038/nrmicro.2017.172

Risks SCoEaNIH. (2015). *Opinion on Synthetic Biology II. Risk Assessment Methodologies and Safety Aspects.* European Commission.

Rodrigues, J. L., & Rodrigues, L. R. (2018). Potential Applications of the Escherichia coli Heat Shock Response in Synthetic Biology. *Trends in Biotechnology*, *36*(2), 186–198. https://doi.org/10.1016/j.tibtech.2017.10.014

Rolfe, M. D., Rice, C. J., Lucchini, S., Pin, C., Thompson, A., Cameron, A. D. S., … Hinton, J. C. D. (2012). Lag Phase Is a Distinct Growth Phase That Prepares Bacteria for Exponential Growth and Involves Transient Metal Accumulation. *Journal of Bacteriology*, *194*(3), 686–701. https://doi.org/10.1128/JB.06112-11

Ronchel, M. C., & Ramos, J. L. (2001). Dual System to Reinforce Biological Containment of Recombinant

Bacteria Designed for Rhizoremediation. *Applied and Environmental Microbiology*, *67*(6), 2649–2656. https://doi.org/10.1128/AEM.67.6.2649-2656.2001

Roth, J. R. (1978). *Tandem Genetic Duplications in Salmonella typhimurium : Amplification of the Histidine Operon*. 53–71.

Rousk, J., Bååth, E., Brookes, P. C., Lauber, C. L., Lozupone, C., Caporaso, J. G., … Fierer, N. (2010). Soil bacterial and fungal communities across a pH gradient in an arable soil. *ISME Journal*, *4*(10), 1340–1351. https://doi.org/10.1038/ismej.2010.58

Rovner, A. J., Haimovich, A. D., Katz, S. R., Li, Z., Grome, M. W., Gassaway, B. M., … Isaacs, F. J. (2015). Recoded organisms engineered to depend on synthetic amino acids. *Nature*, *518*(7537), 89–93. https://doi.org/10.1038/nature14095

Santos, V. E., Casas, J. A., & Go, E. (2000). *Xanthan gum: production, recovery, and properties*. *18*. https://doi.org/10.1016/S0734-9750(00)00050-1

Sarai, A., & Takeda, Y. (1989). Lambda repressor recognizes the approximately 2-fold symmetric half-operator sequences asymmetrically. *Proceedings of the National Academy of Sciences of the United States of America*, *86*(17), 6513–6517. https://doi.org/10.1073/pnas.86.17.6513

Sarwat, F., Qader, S. A. U., Aman, A., & Ahmed, N. (2008). Production & characterization of a unique dextran from an indigenous Leuconostoc mesenteroides CMG713. *International Journal of Biological Sciences*, *4*(6), 379–386. https://doi.org/10.7150/ijbs.4.379

Sauer, R. T., Ross, M. J., & Ptashne, M. (1982). Cleavage of the lambda and {P}22 repressors by {recA} protein. *Journal of Biological Chemistry*, *257*(8), 4458–4462. Retrieved from http://www.jbc.org.ezp-prod1.hul.harvard.edu/content/257/8/4458

Sawitzke, J. A., Thomason, L. C., Costantino, N., Bubunenko, M., Datta, S., & Court, D. L. (2007). Recombineering: In Vivo Genetic Engineering in E. coli, S. enterica, and Beyond. *Methods in Enzymology*. https://doi.org/10.1016/S0076-6879(06)21015-2

Sayler, G., Biotechnology, S. R.-C. O. in, & 2000, U. (2000). Field applications of genetically engineered microorganisms forbioremediation processes. *Elsevier*, *11*, 286–289. Retrieved from https://www.sciencedirect.com/science/article/pii/S0958166900000975

Schmidt, M., & De Lorenzo, V. (2000). Synthetic bugs on the loose: Containment options for deeply engineered (micro)organisms. *New Scientist*, *165*(2221), 18. https://doi.org/10.1016/j.copbio.2016.01.006

Schmidt, M., & De Lorenzo, V. (2012). Synthetic constructs in/for the environment: Managing the interplay between natural and engineered Biology. *FEBS Letters*, Vol. 586, pp. 2199–2206. https://doi.org/10.1016/j.febslet.2012.02.022

Šeputiene, V., Sužiedelis, K., Normark, S., Melefors, Ö., & Sužiedeliene, E. (2004). Transcriptional analysis of the acid-inducible asr gene in enterobacteria. *Research in Microbiology*, *155*(7), 535–542. https://doi.org/10.1016/j.resmic.2004.03.010

Seyed Mohammad Kazem Hosseini Asl, S. D. H. (2000). Determination of the Mean Daily Stool Weight, Frequency of Defecation and Bowel Transit Time: Assessment of 1000 Healthy Subjects. *Arch. Iranian Med.*, *3*, 101–115. Retrieved from http://ams.ac.ir/AIM/0034/asl0034.html

Shultzaberger, R K, Bucheimer, R. E., Rudd, K. E., & Schneider, T. D. (2001). Anatomy of Escherichia coli ribosome binding sites. *Journal of Molecular Biology*, *313*, 215–228. https://doi.org/10.1006/jmbi.2001.5040

Shultzaberger, Ryan K., Chen, Z., Lewis, K. A., & Schneider, T. D. (2007). Anatomy of Escherichia coli ??70 promoters. *Nucleic Acids Research*, *35*(3), 771–788. https://doi.org/10.1093/nar/gkl956

Slechta, E. S., Bunny, K. L., Kugelberg, E., Kofoid, E., Andersson, D. I., & Roth, J. R. (2003). Adaptive mutation: General mutagenesis is not a programmed response to stress but results from rare coamplification of dinB with lac. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.1735464100

Sleight, S. C., Bartley, B. a, Lieviant, J. a, & Sauro, H. M. (2010). Designing and engineering evolutionary robust genetic circuits. *Journal of Biological Engineering*, *4*(1), 12. https://doi.org/10.1186/1754-1611-4-12

Steidler, L, Neirynck, S., Huyghebaert, N., Snoeck, V., Vermeire, A., Goddeeris, B., … Remaut, E. (2003). Biological containment of genetically modified Lactococcus lactis for intestinal delivery of human interleukin 10. *Nat Biotechnol*, *21*(7), 785–789. https://doi.org/10.1038/nbt840

Steidler, Lothar. (2003). Genetically engineered probiotics. *Bailliere's Best Practice and Research in Clinical Gastroenterology*, *17*(5), 861–876. https://doi.org/10.1016/S1521-6918(03)00072-6

Stirling, A. (2007). Risk, precaution and science: towards a more constructive policy debate. Talking point on the precautionary principle. *EMBO Reports*, *8*(4), 309–315.

Stirling, F., Bitzan, L., O'Keefe, S., Redfield, E., Oliver, J. W. K., Way, J., & Silver, P. A. (2017). Rational Design of Evolutionarily Stable Microbial Kill Switches. *Molecular Cell*, *68*(4), 686-697.e3. https://doi.org/10.1016/j.molcel.2017.10.033

Stirling, F., Naydich, A., Bramante, J., Barocio, R., Certo, M., Wellington, H., … Silver, P. (2019). Synthetic cassettes for pH-mediated sensing, counting and containment. *BioRxiv*, 740902. https://doi.org/10.1101/740902

Strand, T. A., Lale, R., Degnes, K. F., Lando, M., & Valla, S. (2014). A new and improved host-independent plasmid system for RK2-based conjugal transfer. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0090372

Subsoontorn, P., & Endy, D. (2012). Design and {Analysis} of {Genetically} {Encoded} {Counters}. *Procedia Computer Science*, *11*, 43–54. https://doi.org/10.1016/j.procs.2012.09.006

Sundarram, A., & Murthy, T. P. K. (2014). α -Amylase Production and Applications : A Review. *Journal of Applied & Environmental Microbiology*, *2*(4), 166–175. https://doi.org/10.12691/jaem-2-4-10

SUZIEDELIENE E. (1999). *The acid inducible asr gene in E. coli: trancriptional control by the phoBR operon*. *181*(7), 2084.

Takeda, Y., Sarai, a, & Rivera, V. M. (1989). Analysis of the sequence-specific interactions between Cro repressor and operator DNA by systematic base substitution experiments. *Proceedings of the National Academy of Sciences of the United States of America*, *86*(January), 439–443. https://doi.org/10.1073/pnas.86.2.439

Thomason, L. C., Costantino, N., & Court, D. L. (2007). *E. coli* Genome Manipulation by P1 Transduction.

Current Protocols in Molecular Biology, (July), 1.17.1-1.17.8. https://doi.org/10.1002/0471142727.mb0117s79

Thomason, L., Court, D. L., Bubunenko, M., Costantino, N., Wilson, H., Datta, S., & Oppenheim, A. (2007). Recombineering: genetic engineering in bacteria using homologous recombination. *Current Protocols in Molecular Biology / Edited by Frederick M. Ausubel ... [et Al.]*, *Chapter 1*, Unit 1.16. https://doi.org/10.1002/0471142727.mb0116s78

Toprak, E., Veres, A., Yildiz, S., Pedraza, J. M., Chait, R., Paulsson, J., & Kishony, R. (2013). Building a morbidostat: An automated continuous-culture device for studying bacterial drug resistance under dynamically sustained drug inhibition. *Nature Protocols*, *8*(3), 555–567. https://doi.org/10.1038/nprot.2013.021

Torres, B., Jaenecke, S., Timmis, K. N., García, J. L., & Díaz, E. (2000). A gene containment strategy based on a restriction-modification system. *Environmental Microbiology*, *2*(5), 555–563. https://doi.org/10.1046/j.1462-2920.2000.00138.x

Torres, B., Jaenecke, S., Timmis, K. N., García, J. L., & Díaz, E. (2003). A dual lethal system to enhance containment of recombinant micro-organisms. *Microbiology*, *149*(12), 3595–3601. https://doi.org/10.1099/mic.0.26618-0

Torres, L., Krüger, A., Csibra, E., Gianni, E., & Pinheiro, V. B. (2016). Synthetic biology approaches to biological containment: Pre-emptively tackling potential risks. *Essays in Biochemistry*, *60*(4), 393–410. https://doi.org/10.1042/EBC20160013

Van Pel, A., & Colson, C. (1974). DNA restriction and modification systems in Salmonella. *Molecular and General Genetics MGG*. https://doi.org/10.1007/bf00433901

Wang, C., Zhao, T., Li, Y., Huang, G., White, M. A., & Gao, J. (2017). Investigation of endosome and lysosome biology by ultra pH-sensitive nanoprobes. *Advanced Drug Delivery Reviews*, *113*, 87–96. https://doi.org/10.1016/j.addr.2016.08.014

Wang, H. H., Isaacs, F. J., Carr, P. A., Sun, Z. Z., Xu, G., Forest, C. R., & Church, G. M. (2009). Programming cells by multiplex genome engineering and accelerated evolution. *Nature*. https://doi.org/10.1038/nature08187

Wang, K., Fredens, J., Brunner, S. F., Kim, S. H., Chia, T., & Chin, J. W. (2016). Defining synonymous codon compression schemes by genome recoding. *Nature*, *539*(7627), 59–64. https://doi.org/10.1038/nature20124

Way, J. C., Silver, P. a, & Howard, R. J. (2011). Sun-driven microbial synthesis of chemicals in space. *International Journal of Astrobiology*, *10*(04), 359–364. https://doi.org/10.1017/S1473550411000218

Wright, O., Delmans, M., Stan, G. B., & Ellis, T. (2015). GeneGuard: A modular plasmid system designed for biosafety. *ACS Synthetic Biology*, *4*(3), 307–316. https://doi.org/10.1021/sb500234s

Wright, O., Stan, G. B., & Ellis, T. (2013). Building-in biosafety for synthetic biology. *Microbiology (United Kingdom)*, *159*(PART7), 1221–1235. https://doi.org/10.1099/mic.0.066308-0

Yamaguchi, Y., & Inouye, M. (2011). Regulation of growth and death in Escherichia coli by toxin-antitoxin systems. *Nat Rev Microbiol*, *9*(11), 779–790. https://doi.org/10.1038/nrmicro2651

Yamanaka, K. (1999). Cold Shock Response in Escherichia coli JMMB Symposium. *J. Mol. Microbiol. Biotechnol*, *1*(2), 193–202.

Yansura, D. G., & Henner, D. J. (1990). Use of Escherichia coli trp promoter for direct expression of proteins. *Methods in Enzymology*, *185*(1981), 54–60. https://doi.org/10.1016/0076-6879(90)85007-B

Yoneda, Y. (1980). Increased production of extracellular enzymes by the synergistic effect of genes introduced into Bacillus subtilis by stepwise transformation. *Applied and Environmental Microbiology*, *39*(1), 274–276.

Yoshida, S., Hiraga, K., Takehana, T., Taniguchi, I., Yamaji, H., Maeda, Y., … Oda, K. (2016). A bacterium that degrades and assimilates poly(ethylene terephthalate). *Science*, *351*(6278), 1196–1199. https://doi.org/10.1126/science.aad6359

Yu, D., Ellis, H. M., Lee, E. C., Jenkins, N. A., Copeland, N. G., & Court, D. L. (2000). An efficient recombination system for chromosome engineering in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.100127597

Zaslaver, A., Bren, A., Ronen, M., Itzkovitz, S., Kikoin, I., Shavit, S., … Alon, U. (2006). A comprehensive library of fluorescent transcriptional reporters for Escherichia coli. *Nature Methods*. https://doi.org/10.1038/nmeth895

Zheng, J. H., Nguyen, V. H., Jiang, S. N., Park, S. H., Tan, W., Hong, S. H., … Min, J. J. (2017). Two-step enhanced cancer immunotherapy with engineered Salmonella typhimurium secreting heterologous flagellin. *Science Translational Medicine*. https://doi.org/10.1126/scitranslmed.aak9537

Zheng, Y., Lajoie, M. J., Italia, J. S., Chin, M. A., Church, G. M., & Chatterjee, A. (2016). Performance of optimized noncanonical amino acid mutagenesis systems in the absence of release factor 1. *Mol. BioSyst.*, *12*(6), 1746–1749. https://doi.org/10.1039/C6MB00070C

Summary Material

Chapter 2 supplementary Information

Large-scale recoding of a bacterial genome by iterative recombineering of synthetic DNA

Yu Heng Lau[a,b], Finn Stirling[a,b], James Kuo[a,b], Michiel A. P. Karrenbelt[a,c], Adam Riesselman[d], Connor A. Horton[a,b], Elena Schaefer[a,b], Yujia A. Chan[a,b], David Lips[a,b], Matthew T. Weinstock[e], Daniel G. Gibson[e,f], Jeffrey C. Way[a,b] and Pamela A. Silver[a,b]

[a] Wyss Institute for Biologically Inspired Engineering, Harvard University, 3 Blackfan Circle, 5th Floor, Boston, MA 02115, USA

[b] Department of Systems Biology, Harvard Medical School, 200 Longwood Avenue, Alpert 536, Boston, MA 02115, USA

[c] Systems and Synthetic Biology, Wageningen University, Dreijenplein 10, 6703 HB Wageningen, The Netherlands

[d] Program in Biomedical Informatics, Harvard Medical School, Boston, MA 02115, USA

[e] Synthetic Genomics, Inc., 11149 North Torrey Pines Road, La Jolla, CA 92037, USA

[f] Synthetic Biology and Bioenergy Group, J. Craig Venter Institute, 4120 Capricorn Lane, La Jolla, CA 92037, USA

SI 1. General experimental procedures

Yeast amino acid dropout mixes and yeast nitrogen base were obtained from US Biological. LB-agar mix was obtained from EMD Millipore. All other media components were obtained from BD Difco. Unless otherwise specified, all chemical reagents were obtained from Sigma-Aldrich.

DNA electrophoresis gels were run with 1% agarose (SeaKem) in TAE buffer and stained with GelRed dye (Biotium), using a 1 kb plus ladder (Thermo Fisher) for sizing.

Antibiotics were used at the following concentrations when selecting for genomic integration: 5 µg/mL chloramphenicol, 50 µg/mL kanamycin, 10 µg/mL gentamicin, 50 µg/mL carbenicillin, 50 µg/mL spectinomycin, 7.5 µg/mL tetracycline.

SI 1.1 Media formulations

Lysogeny broth (LB): 0.5% bacto yeast extract, 1% bacto tryptone, 1% NaCl

SOC medium: 0.5% bacto yeast extract, 2% bacto tryptone, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM D-glucose

YPAD medium: 1% bacto yeast extract, 2% bacto peptone, adenine hemisulfate, 2% D-glucose

SD-Trp medium: 0.192% Trp dropout mix, 0.67% yeast nitrogen base with ammonium sulfate, 0.008% adenine hemisulfate, 2% D-glucose

Plates formulations were made with the addition of 2% bacto agar.

SI 1.2 Buffer formulations

Tris-EDTA buffer: 10 mM Tris, 1 mM Na_2EDTA, pH 8.0

TAE buffer: 40 mM Tris, 20 mM acetic acid, 1 mM Na_2EDTA, pH 8.3

SI 1.3 Reagents for DNA assembly in yeast

SPE buffer: 1 M sorbitol, 10 mM Na_2EDTA, 10 mM sodium phosphate, pH 7.5

STC buffer: 1 M sorbitol, 10 mM $CaCl_2$, 10 mM Tris, pH 7.5

20% PEG 8000 solution: 20% PEG 8000, 10 mM Tris, 10 mM $CaCl_2$, pH 7.5

SOS medium: 1 M sorbitol, 0.25 % bacto yeast extract, 0.5 % bacto peptone, 6 mM $CaCl_2$

Top agar -Trp: 1 M sorbitol, 0.192% Trp dropout mix, 0.67% yeast nitrogen base with ammonium sulfate, 0.008% adenine hemisulfate, 2% D-glucose, 3% bacto agar

SD-Trp plates: identical to top agar except 2% bacto agar

SI 2. Computational design of the recoded *Salmonella* genome

All recoding changes, LguI site changes and genomic deletions are annotated in the GenBank file for the recoded *Salmonella* genome (Recoded Salmonella genome.gb, SI file 3).

SI 2.1 Reference *Salmonella* genome and annotations used

The recoded *Salmonella* genome was based on the sequence published by McClelland and coworkers (Michael McClelland et al., 2001). Annotations were obtained from the reference sequence NC_003197 (version NC_003197.1). Initial design decisions were based on the older GenBank file AE006468 (version AE006468.1), with some DNA constructs used in this project corresponding to the open reading frames as defined by this set of annotations. We also note that a recent updated reference sequence (version NC_003197.2) was made available on NCBI near the time of completion of this manuscript.

SI 2.2 Rationale for choice of leucine codons for recoding

Frequency of genome-wide codon occurrence in *Salmonella* for the six-codon amino acids was determined (based on AE006468.1). Removing the tRNA components for the 'two-group' codons should not interfere with the remaining four codons. Arginine was not chosen for recoding due to the low frequency of the occurrence of AGA and AGG codons, which would also be uncommon in incoming DNA by horizontal gene transfer, thus diminishing the efficiency of genetic isolation. Leucine was chosen over serine as there were less target codons (TTA and TTG) present in overlapping reading frames (97 vs 148 in AE006468.1). In instances where TTG serves as an alternative start codon and thus does not encode for leucine, this codon was not altered in the recoded genome design.

This approach differs from that previously taken by both Church and co-workers(Ostrov et al., 2016) and Chin and co-workers(K. Wang et al., 2016). In these previous examples, all overlapping gene pairs were refactored to separate the two genes.

Table 2.2.1. Number of each codon present in the *Salmonella* genome (annotations from AE006468.1) for leucine, serine and arginine.

Leucine		**Serine**		**Arginine**	
CTT	15132	TCT	9164	CGT	24581
CTC	13742	TCC	13087	CGC	30885
CTA	6467	TCA	7545	CGA	4508
CTG	70584	TCG	12574	CGG	8608
TTA	17094	AGT	8907	AGA	2656
TTG	16395	AGC	22641	AGG	1783

A full list of codon changes can be found in SI file 2 (Tab 5: all codon changes).

SI 2.3 Treatment of changes in overlapping reading frames

When a TTA or TTG codon was present in two overlapping reading frames, mutations were chosen that resulted in a synonymous mutation in both reading frames where possible. If a non-synonymous mutation was required, all possible nucleotide permutations were translated into their mutated protein products in both reading frames, which were then analyzed by PROVEAN to assess their predicted impact on protein function (Choi et al., 2012). The following table lists all 49 protein mutations in the recoded genome design (from NC_003197.1), chosen for minimal impact according to PROVEAN.

Table 2.3.1. Mutations introduced by recoding in overlapping genes. The mutations were chosen to minimize impact on protein function as estimated by PROVEAN. In some cases, the leucine codon is recoded, whilst the leucine is mutated to a different residue in other cases.

Gene	Protein mutations	Original protein sequence
STM0027	L8S	MPPQPSFLPMNKFFLRCAIYWCLLPISWAQAGVVIGGTRFIYHAGAPALSVPVSNRSEASWLIDTHILPGGRWPGTKNEGNIMPFVVTPPLFMLSARQENSMRVVYT GAPLPADRESLFTLSIAAIPSGKPEANRVQMAFRSALKLLYRPDGLAGNPQQAYRHLIWSLTPDGATVRNPTPYYVTLFLLRANERAQDNAGVVAPFATRQTDWCRHT VRCTVRWQSINDYGRVMPAQTVDLTRIH*
STM0333	L312S	MKTTPVGKKHRQPLEGWPSVEDLFVFVTVARNGGFARAALELGLSPSYISKRIAILEKCLNARLFFRNNRVMRLTPEGENALGGAMQVVSEMDGFVSRLDNQRGVLA GNITINCSFGFGHKYVAEALSSFMIAYPDITVKLTLTDREVDLVEEGIDIEIRVGDDINELYIARQLATNRRILCASPEYLERHGKPESVSALVQHDCLMIQEKNSAFGNWIL TDGKQQTHCRLNGFHSSNSGSVVLIWALKGHGITLRSEWDVAQYIERGELVRVLPQWYQEANIWAVYTRRSSSSDRIKICIDFLTEHLAQCLPGGKAPGVL*
STM0336	L8S	MRNYRQWLVFSKVILTLLGLTGWYGPAQAAVNIDRTRIIFASDEVAQSLTLSNDNTTPMLLQVWTDAGNIDASPDNSKTPLVALPPVFKMQPGELRTLRLLLSSRQQL ATDRESLFWLNIYQIPPVTQDIKNHPRKLVLPLRLRLKILIRPTGLKAPTEAEEKKLRFIAKENTIRIVNPTSWYMSLTLTTDDKKSIGDIMVAPKTARDVPLTRSPTPGASIN YAVISDSGNWRQYTAILEKYISENDFTPEF*
STM0351	V1048A	MKFTHFFIARPIFAIVLSLLMLLAGAIAFLKLPLSEYPAVTPPTVQVSASYPGANPQVIADTVAAPLEQVINGVDGMLYMNTQMAIDGRMVISIAFEQGTDPDMAQIQV QNRVSRALPRLPEEVQRIGVVTEKTSPDMLMVVHLVSPQKRYDSLYLSNFAIRQVRDELARLPGVGDVLVWGAGEYAMRVWLDPAKIANRGLTASDIVTALREQNVQ VAAGSVGQQPEASAAFQMTVNTLGRLTSEEQFGEIVVKIGADGEVTRLRDVARVTLGADAYTLRSLLNGEAAPALQIIQSPGANAIDVSNAIRGKMDELQQNFPQDIE YRIAYDPTVFVRASLQSVAITLLEALVLVVLVVVMFLQTWRASIIPLVAVPVSLVGTFALMHLFGFSLNTLSLFGLVLSIGIVVDDAIVVVENVERHISQGKSPGEAAKKAM DEVTGPILSITSVLTAVFIPSAFLAGLQGEFYRQFALTIAISTILSAINSLTLSPALAAILLRPHHDTAKADWLTRLMGTVTGGFFHRFNRFFDSASNRYVSAVRRAVRGSVIV MVLYAGFVGLTWLGFHQVPNGFVPAQDKYYLVGIAQLPSGASLDRTEAVVKQMSAIALAEPGVESVVVFPGLSVNGPVNVPNSALMFAMLKPFDEREDPSLSANAIA GKLMHKFSHIPDGFIGIFPPPPVPGLGATGGFKLQIEDRAELGFEAMTKVQSEIMSKAMQTPELANMLASFQTNAPQLQVDIDRVKAKSMGVSLTDIFETLQINLGSLY VNDFNRFGRAWRVMAQADAPFRMQQEDIGLLKVRNAKGEMIPLSAFVTIMRQSGPDRIIHYNGFPSVDISGGPAPGFSSGQATDAIEKIVRETLPEGMVFEWTDLVY QEKQAGNSALAIFALAVLLAFLILAAQYNSWSLPFAVLLIAPMSLLSAIVGVWVSGGDNNIFTQIGFVVLVGLAAKNAILIVEFARAKEHDGADPLTAVLEASRLRLRPILM TSFAFIAGVVPLVLATGAGAEMRHAMGIAVFAGMLGVTLFGLLLTPVFYVVVRRMALKRENRVDSHDQQA*
STM0420	L4S	MTILPRHKDVAKSRLNLRNPWHLLATGFGSGLSPVVPGTMGSLAAIPFWYLMTFLPWQLYSLVVMFGICIGVYLCHQTAKDMGVHDHGSIVWDEFIGMWITLMALP TTDWQWVAAGFVIFRILDMWKPWPIRWFDRNVHGGMGIMIDDIVAGVISAGILYFIGHHWPLGII*
STM0577	I146T	MRHVYVASHGPFARGLINSLSLLIGDEHGVTPVCAYDGDIVTTEQLEQTLENLIAQANGEEVVVFTDLLGGSINNSAAKVLMRHRHVFVVAGVNMTLLLEFLLCEEEST DAAITYATNAARESIVFINTLITQPSADLQGESHDQISAH*
STM0625	I548T, I551T	MSCQKNKKLFPFFRQLAFPLRIFLLILVVSVFIVAALAQYLSASFEDYLASHVRDMAMNQAKIIASNDSIIAAVKNRDYKRLAIIANKLQRGTDFDYVVIGDRHSIRLYHPNP EKIGYPMQFTKPGALERGESYFITGKGSIGMAMRAKTPIFDNEGNVIGVVSIGYLVSKIDSWRLDFLLPMAGVFVLLLVVLMLLSWFFAAHIRRQMLGMEPKQIARVVR QQEALFSSVYEGLIAVDPEGHITAINRNARKMLGLPSPGRQWLGKLIAEVVNPADFFTCQIAERRQDVMANFNGLSVIANREAIRSGEELLGAIISFRSKDEIATLNAQLT QIKQYVESLRTLRHEHLNWMSTLNGLLQMKEYDRVREMVQGESQAQQQLIDSLRGAFADRQVAGLLFGKVQRARELGLKMVIVPGSQLHQLPEGLDSTEFAAIVGN LLDNAFEASLRTQEGDKTIELFLSDEGDEVVIEVADQGCGVPEALREKIFEQGVSTRTDEPGEHGIGLYLIASYVGRCGGVITLEDNDPCGTLFSLFLPKVKKNDDRTINPID R*
STM0953	I7T	MAKEDNIEMQGTVLETLPNTMFRVELENGHVVTAHISGKMRKNYIRILTGDKVTVELTPYDLSKGRIVFRSR*
STM0993	L3S	MSLTNIEQVMPVKLAQALANPLFPALDSALRSGRHIGLDELDNHAFLMDFQEYLEEFYARYNVELIRAPEGFFYLRPRSTTLIPRSVLSELDMMVGKILCYLYLSPERLANE GIFTQQELYDELLTLADEAKLLKLVNNRSTGSDVDRQKLQEKVRSSLNRLRRLGMVWFMGHDSSKFRITESVFRFGADVRAGDDPREAQRRLIRDGEAMPIENHLQLN DETEESQPDSGEEE*
STM1148	L12S	MKKLPGFTQDYLLSKATTLPDKTRLERAVEPLCARHPGECGILALDNSLDAFAARYRLTEMAARTLDVQYYIWEDDMSGRLLFSVLLSAAKRGVHVRLLLDDNNTPGLD DTLRLLDSHPNIEVRLFNPFSFRTLRALGYLTDFARLNRRMHNKSYTADGVVTLVGGRNIGDAYFGAGEEPLFSDLDVMAIGPVVNDVANDFERYWRCSSVSTLQQVLS LSEQELTQRIELPESWYNDEITRRYLHKLETSQFMADLDRGTLPLIWAKTRLLSDDPSKGEGKAQRHSLLPQRLFDVMGSPTERIDIISAYFVPTRAGVAQLLNLVRKGVKI AILTNSLAANDVAVVHAGYARWRKKLLRYGVELYELKPTREHETAVHDRGLTGNSGSSLHAKTFSIDGSKVFIGSLNFDPRSTLLNTEMGFVIESETLATLIHKRFTQSQR DAAWQLRLDRWGRINWIDRQQEEEKVLKKEPATRFWQRVLVRLAAILPVEWLL*

STM1301	L129S, T135A	MIKKLDVVAAIIECDGKILLAQRPAHADQAGLWEFAGGKVEPGETQPQALIRELREELGIDATPGVYIASHQRDVSGRRIHLHAWHVPAFNGLIRALEHQALAWCTPEE ALEYPLAPADIPLLQAFIALRDARLTDSC*
STM1371	T245S	MLSIKDLQVSVEEKAILRGLNLEIRPGEVHAIMGPNGSGKSTLSATLAGREDYEVTGGSVTFKGKDLLELSPEERAGEGIFMAFQYPVEIPGVSNQFFLQTALNAVRAYRG QASLDRFDFQDLMEEKIALLKMPEDLLTRSVNVGFSGGEKKRNDILQMAVLEPELCILDESDSGLDIDALKIVAEGVNALRDDKRAFIIVTHYQRILDYIKPDYVHVLYQG RIVRSGDFTLVKQLEEQGYGWLTEQQ*
STM1372	L4F	MAGLPNSSKALQQWQHLFEEKGESRTEQARQHLQQMLRLGLPTRKHEDWKYTPLEGLTHSQFIQQCATISAAQRDALALQIDAVRLVFVDGRFMPELSDSTQNSGF DVSVRDERQALAAPVQPEVFLHLTESLAQCVTYIQVRRRNQRPTRPLLLMHITQGVDGDELNTAHYRHHLALAEGAEATVIEHYVSLTAAKHFTGARLTMNVADNAQLR HFKLAFENASSYHFAHNDLLLATDASAFSHSFLLGAAVLRHHSSSQLNGENATLRLNSLAMPVKNEVCDTRTWLEHNKGYCNSRQLHKTIVSDKGRAVFNGLINVAQH AIKTDGQMTNNNLLLGKLAEVDTKPQLEIYADDVKCSHGATIGRIDDEQMFYLQSRGIRQQEARHMILYAFAAELTEAIHDSALKQQVLARIGQRLPGGLV*
STM1441	L674F	MKLSLPALRNTPWFKATSGQWRYALRNTIAMCLALTFAYYLNLDEPYWAMTSAAVVSFPTVGGVISKSLGRIAGSLLGATAALIIAGHTLNEPWLFLFSMAAWIGFCT WACAHFTNNAAYAFQLSGYTAAIIAFPMVNIVEITQLWDIAQARVCEVIVGILCGGMMMMILPSTSDGTALLTAFKNMHARLLEHASLLWQPETTDAIRSAHEGVIG QILTMNLLRIQAFWSHYRFFRRQNALLNALLHQQLRLTSVISSLRRMLLNWPTPPENSREVIEQLLAELAKPRADSYTVAQIIAPLRPQDEQDYRHLAFWQRLRYFCQLYL RSSRQLYLIESGAPVDQIHIRRTPGLARHTDNAEAIWSGVRTFCTLTVIGAWSIGAQWESGPGALTLAAISCVLYSIVATPFKSLSLLMRTLVLLSLFSFVVKFGLMVQITDL WQFLLFLFPLFVTMQLLKLQMPKLAGLWGQLIVFMGSFIAVTNPPVYDFADFLNDNTAKIVGVAISWLAFAILRPGSDAVKSRRHIRALRRDFVDQLSRHPSHNESEFE SLTYHHVSQLSNSQDALARRWLLRWGVVLLNCSHVVWQLRAWESRSDPLSRVRDICISLLRDVMSERGVQQRPLAVTLQELQRICDTLAHHHQPAAHELAAIIWRLH CSLSQLEQAPAQGTLAPGYLMTPQA*
STM1442	F6S	MSGPRFTDRVVNMSLKTIKYFSTLIVAVVAVLAAWWLWNYYMQSPWTRDGKIRAEQVSVTPQVSGSITQLNIKDNQFVNAGDVLFVIDKTPFHIAELNAQAQLAKA QSDLAKANNEADRRRHLSRNYISAEDLDSANLNVKAMQANVDVALATLKQAQWQLSQTEVKAPVSGWVTNLSTRTGDYASTGKPLFALVDSHSFYVMGYFEETKLR HIREGEPALITLYSGNVKLQGHVGSIGRAIYDQSVESDSGLVPDIKPNVPWVRLAQRVPVRIEFDALPQDITLVSGTTCSVAIGQR*
STM1528	L12E, L20F, K44R	MGITIMQKTSLLFSAAAITLTLMASSASAATPTAAQNEATSTVKQEITEGINRYLYSIDKADPTLGKQLFYVSPETSFIHPRGHERGWSQIAENFYGTTMGKTFSKRTLKLD APPAIHVYGNAAVAEFDWHFTAVRRDNGQTQHTTGRESQVWAKIPNTGWRIVHVHYSGPAKTGVGEGY*
STM1529	N40S	MPSVISCLTVDVASFCAAVGVAADADDAIRVNVIAAAENNSDVFCIIVIPIWSVFETLYLSCYVHLYTKGAWLTNNKKHPSIKLPRKGSSAIKNIVFVCNGMTQNINI*
STM1537	L2S, L3S	MLLIIMEGSWSGFRYDDDDSEVSMHLKEITSQGYYIYEAPVRLWHWITALSIVVLAVTGYFIGRPLPSIQGEATFMFWMGWIRLIHFTTAYIFTVALLFRIYWACVGNEY AREMFLVPFWRRAWRKGVISEIRWYFFLEKEAHRYYGHNPVAGLAVMFYFWMSVLMVCSGFALYGEGLGTDSWAYQWFGWMIRLTGNDSLALHFWHRLGMWF IIAFVIAHVYTAIREDIMSRQSVISVMISGWRWFR*
STM1929	F263S	MAEPLTVSPELTANYAYFFDLDGTLAEIKPHPDQVVVPHKILQLLDRLAAHNAGALALISGRSMTELDALAKPFRFPLAGVHGAERRDINGKTHIVRLPEAVVREVEALLR STLVALPGTELESKGMAFALHYRQAPEHEAALLALAQHVTQHWPQLALQPGKCVVEIKPKGTNKGEAIAAFMQEAPFAGRIPVFVGDDLTDEAGFGVVNHAGGISVK VGVGATQAAWRLESVPDVWRWLEQINYPQQEQQVMNNRRDGYESFSRSI*
STM2162	K243E	MKRLCDPLLWLIVLFLLLLFGLPYSQPFFAALFPDLPRPVYQQESFAALALAHFWLVGISSLFAVVVGVGAGIAVTRESGKEFRPLVETIAAVGQTFPPVAVLAIAVPVMG FGQQPAIIALILYGVLPILQATLAGLGAVPASVMSVASGMGMSRRQQLYQVELPLAAPVILAGIRTSVIINIGTATIASTVGASTLGTPIIIGLSGFNTAYVIQGALLVALAAII IDRLFERLTRVLTRHAK*
STM2397	V667A	MKGRLLQRLRQLSISNSLRGAFLTGALLTLIVSMVSLYSWHEQSSQVRYSLDEYFPRIHSAFLIEGNLNLAVDQLNEFLLAPNTTVRLQLRTQIIQHLDKIERLSQGLQLAER RQLAVILQDSRTLLAELDNALYNMFLVREKVSELSARIDWLHDDFTTELNSLVQDFTWQQGTLLDQIEANQGDAAQYLQRSREVQNEQQQVYTLARIENQIVDDLRD RLNELKSGNNDGMLVETHIRYLENLKKTADENIRALDDWPSTITLRQTIDELLEIGMVKNKMPDTMRDYVAAQKALLDASRAREATLGRFRTLLEAQLGSSHQQMQTF NQRLEQIVRVSGGLILVATLLALLLAWGLNHYFIRSRLVKRFTALNQAVVQIGLGRTDSTIPVYGRDELGRIARLLRHTLGQLNMQRRQLEQEVAERKEIEADLRAMQDE LIQTAKLAVVGQTMTTLAHEINQPLNALSMYLFTAGRAIEQGQSGQARNTLTKAEGLINRIDAIIRSLRQFTRRAELETPLYPVDLRQTFVAAWELLAMRHQSRQGALSL PTDTVWVSGDEVRIQQVLVNVLANALDACSHDAVIAVTWQTQGEALEVYIADNGPGWPVALLPSLLKPFTTSKAVGLGIGLSISVSLMAQMKGDLRLASTLTRNACV VLQFSVTDVDDVE*
STM2928	I347T	MTEFDNLTWLHGKPQGSGLLKANPEDFVVVEDLGFTPDGEGEHILLRILKNGCNTRFVADALAKFLKIHAREVSFAGQKDKHAVTEQWLCARVPGKEMPDFSAFQLE GCKVLEYARHKRKLRLGALKGNAFTLVLREISDRRDVETRLQAIRDGGVPNYFGAQRFGIGGSNLQGALRWAQSNAPVRDRNKRSFWLSAARSALFNQIVHQRLKKPD FNQVVDGDALQLAGRGSWFVATSEELPELQRRVDEKELMITASLPGSGEWGTQRAALAFEQDAIAQETVLQSLLLREKVEASRRAMLLYPQQLSWNWWDDVTVELR FWLPAGSFATSVVRELINTMGDYAHIAE*
STM3073	L234S	MHPFTSLTLWALAACTTLLLPAQTVLPVYSAAAFLCLLALKSTRRRAKYVAWLMLSLGFGLWLVHGGWLTEWISGQPRDPQRWVYAVTLWLRLLAIVSTSQLWMQY VPVQRFIRALFASRLPPGIAYLFAGPLLVVEQLKRQLTIVHEAQRARGVPLDEGWYQRLRAMPALIVPLTQNALNDLTIRGAALDMRGFRLHRARTTLWAPKDSVLQRV ARYGMVLLILAEAGVWIWLR*
STM3276	L334S	MTDKTIPFSVLDLAPIPEGSSAKEAFTHSLDLARLAEKRGYHRYWLAEHHNMTGIASAATSVLIGYLAANTTTLHLGSGGVMLPNHSPLVIAEQFGTLNTLYPGRIDLGLG RAPGSDQSTMRALRRHMSGDIDNFPRDVAELVDWFDARDPNPHVRPVPGYGEQIPVWLLGSSLYSAQLAAQLGLPFAFASHFAPDMLFQALHLYRTQFKPSARLEK PYAMVCINIIAADSNRDAEFLFTSMQQAFVKLRRGETGQLPPPIENMETFRSPSEQYGVQQALSMSLVGDKAKVRHGLESILRETQADEIMVNGQIFDHQARLHSFDL AMDVKEELLG*
STM3277	T237A	MGRKGLLAIVLLSLFIAFILKFFWLTPYDEDVYLPVEKPVASSLKIIHPGDQLFIRILKAEDKLELWASANNKPYKLYKTWTICAWSGGLGPKHKQGDGKSPEGFYATNKGL LNPNSRYHLAFNIGYPNAYDRANGYTGDFIMVHGNCVSAGCYAMTDAGIEEIYQLVAQALNSGQKSVPVHIFPFTMNDENMRQAQAWPEYNFWRMLKPGYDYFE KNRRLPTITVENRRYKISPTTLP*
STM3399	N180S	MSDTLRPYKNLFPGIGQRVMIDTSSVVIGDVRLADDVGIWPLVVIRGDVNYVAIGARTNIQDGSVLHVTHKSSSNPHGNPLIIGEDVTVGHKVMLHGCTIGNRVLVGM GSIVLDGAIIEDDVMIGAGSLVPQHKRLESGYLYLGSPVKQIRPLSDAERSGLQYSANNYVKWKDDYLSQDNHIQP*
STM3405	F373S	MARTEIWLRLMYVGDLYGEAMLNMANSLIRQPQINRTHLQEAGLTARQAERFLQLPAGVLDETLRWLELPQHHFLCADSEIYPPQLRAIDDYPGAIFIDGDPACLHTC QLAVVGSRSHSWYGERWGRLLCESLAKSGLTITSGLARGIDGVAHNAAVSMGGKSVAVLGNGLAKIYPRRHAMLAENLIATGGAVVSEFPLSTPPLPQHFPRRNRIISG LSKGVLVIEAALRSGSLVTARCALEQGRDVFALPGPIGSPGSEGTHWLIKQGATLVTTPEDILENLQYGLHWLPTTAENSLYSLNQDEAALPFPELLANVGDEVTPVDVV AERAGQPVPAVVAQLLELELAGWIAAVPGGYVRLRRASHVRRTNVFV*
STM3411	I85V, T86A	MSRYQHKKGQIKDNAIEALLHDPLFRQRVEKNKKGKGSYLRKGKHGNRGNWEASGKKVNHFFTTGLLLIKNGYFVLSADRGFRPITLPL*

STM3489	L117P	MKASRSVLLCFCLLMLTGMRDPFRPPEDRCRIAELSQWRYQGAVRKGERWTGILKDSQQKWRRVEEGQTLENGWTIVRLTAETLTLTTGKNCAPPQWRWLRQGADNEAMDSHNTDSLDARRAGGKSGESDAGG*
STM3572	V197A	MKKPTHSGSGQIRIIGGQWRGRKLPVPDSPGLRPTTDRVRETLFNWLAPVMVDAHCLDCFAGSGALGLEALSRYAAQATLLEMDRAVSQQLQKNLATLKANNARVVNTNTLTFLSQPGTPHHVVFVDPPFRKGLLEETLQLLETQGWLADDALIYVESEVENGLPPVPANWALYREKVAGQVAYRLYQRDAQGENDVD*
STM3799	I360V	MVKILGGVVFKPLIASLMLTSAVVYAKPMPLTAARYAQQLGVGMDVDWARTERGIREFDPLVVRDFKAKGLTHVRIRVAGAPTEARLIHLRKLVEACEYYGVIPIIAYQADAYKTDPSASHEKELINWWSVVARYFGQTSPLLGFDLIYEPADKLNHNMASLNRVYDKTIRLIHAIDPQRMIFVAPRMRAAPEDLSALKLPAQSQNYVLAEWHIFPWGPLKSGGKYPWTSGTAAEKAAIRARINAAVRWQHKTGHASWVGGWAPGESIKMTPVASQFAFARFMACELKKAHIPYAINADTQFYDGEEGAWRPAQESLLNAMIAPECETPGKTPGEGNVKPSAPDAISVTPAAASTPGSAIP*
STM3824	L225F	MSHHIVIVEDEPVTQARLQAYFEQEGYRVSVTDSGAGLRDIMEHEHVSLILLDINLPDENGLMLTRALRERSTVGIILVTGRCDQIDRIVGLEMGADDYVTKPLELRELVVRVKNLLWRIDLARPTPQNASENCYMFSGYCLNVMNHTLEHNGEAIKLTRAEYELLLAFVTNPGKVLHRERLLRMLSARRVETPDLRTIDVLVRRLRHKITPELLVTQHGEGYFLASEVY*
STM3887	K473E	MTSKKARSMAGLPWIAAMAFFMQALDATILNTALPAIAHSLNRSPLAMQSAIISYTLTVAMLIPVSGWLADRFGTRRVFMVAVSLFTLGSLACALSSSLSELVIFRVVQGVGGAMMMPVARLALLRAYPRSELLPVLNFVTMPGLVGPILGPVLGGVLVTWASWHWIFLINIPIGVAGILYARKYMPNFTTPRRKFDMTGFFLFGLSLVLFSSGMELFGEKIVATWIASAIIFCSIVLLLAYIRHARRHPTPLISLSLFKTRTFSVGIAGNLATRLGTGCVPFLMPLMLQVGFGYPALIAGCMMAPTALGSIIAKSTVTQVLRRLGYRKTLVGITVFIGLMIAQFSFQSPAMPIWMLVLPLFILGMAMSTQFTAMNTITLADLTDDNASSGNSVLAVTQQLSISLGVAISAAVLRIYEGFAGTSTVEQFHYTFITMGAITIVSALMFMLLRAKDGNNLIKERHKSKPTHAPSKPE*
STM4008	L5F, K7R	MMAMLRKFSITTLPQALEIARYCTRINRTTKQDHKHPPQTSIAESLCDRFHGALKGWLAFCCTNVVLNVHIKALFWCNIVTVVQPFCTMVRMITPFGGNVKVGTDFALYFYGDTASRIEDLVTTTTDHDQSGRVQVCPLNTF*
STM4115	L3S	MTLFAAPRISCEVVGARDEETARIFNIQRYSLNDGQGIRTVVFFKGCPHTCPWCANPESISPRIETVRRENKCLRCTPCLRDADECPSGAFERIGRDITLDELEREVLKDDIFFRTSGGGVTLSGGEVLMQAPFATRFLQRLRRWGVPCAIETAGDTSASRLLPLARACDEVLFDLKIMDAERAREVINMNLPRVLENLRLLVSEGITVIPRLPLIPGFTLNAENLQCALTLLRSLGIKQVHLLPFHQYGEPKYRLLGKSWMMKDIPAPSVQEIALFREMAEQAGFQVTTGG*
STM4324	L114S	MLDVKSQDISIPEAVVVLCTAPDEATAQDLAAKVLAEKLAACATLLPGATSLYYWEGKLEQEYEVQMILKTTVSHQQALIDCLKSHHPYQTPELLVLPVTHGDTDYLSWLNASLR*
STM4356	N5S	MMDHNMKKNPVSIPHSIWPADDIKRLERDAADAFGLTLYELMLRAGDAAFRVARDSYPDTRHWLVLCGHGNNGGDGYVVARLAQAAGISVTLLAQESDKPLPEEAAQARDAWLNAGGIIHAADIIWPEATDLIIDALLGTGIAQAPRDPVAGLIEQANAHPAPVVAVDIPSGLLAQTGATPGAVISAAHTVTFIALKPGLLTGKARDVTGILHYDALGLEGWLASQTPPLRRFDATQLGQWLTPRRPTSHKGDHGRLAIIGGDQGTAGAIRMAGEAALRTGAGLVRVLTRGENIAPLLTARPELMVHELTPQSLEESLTWADVVVIGPGLGQQEWGKKALQKVENVRKPMLWDADALNLLAINPDKRHNRVITPHPGEAARLLGCSVAEIESDRLLSAQRLVKRYGGVVVLKGAGTIIAAEHHPLAIIDAGNAGMASGGMGDVLSGIIGALLGQKFTPYDAACVGCVAHGAAADLLAARYGARGMLATDLFTTLRRIVNPDVIDVNHDESSNSAT*
STM4462	K41R, L44F, K46R	MSYTLNHASSRQIVRDYTQLCFGIKRLMHDLLIANGSYEFKFILSK*
STM4547	L239S	MLPGCCKNGLIISKTPLIQEGLKGAITGNFPDYKLAYCRTIEELTLLQLRRSNLVIADLAVNNASPRAICEYFYSLLSQYRDIHWVFLVPKSCYPHAVDLLMGPVSTLLSDEEPIENLISVIHAGNARSERISKTLLSPQVPSEIQQSHDRPIVLTLSERKVLRLLGKGWGINQIAALLKKSNKTISAQKNSAMRRLSIHSNAEMYAWINSSQGARELNLPSVYGETMEWKTESAREMLRS*
STM4565	L7S	MNNRCALVSKIIAFSCVDGPGSRLALFLQGCNLRCKNCHNPWTMGRCNDCGECVSQCPHHALSIDGGKVVWRSDVCEQCDTCLKMCPQQATPMAQTMSVDDVLRHIRKASLFIEGITVSGGEATTQLPFIVALFTAIKADPQLQRLTCLVDSNGQLSETGWQKLLPVCDGVMLDLKAWKSECHHRLTGRDNTHIKHSIRFLAARGKLAELRLLVIPGQVDYAAHIDSLAAFITSLGEVPVRLNAFHAHGVYGEAKAWPGATPQEVERLADGLRARGVAKLILPALYL*

Whilst refactoring overlapping genes was intentionally avoided to limit the potential impact on the fitness and stability of the genome, there was one instance where overlapping genes were split apart due to the presence of four leucine codons. Between STM0521 and STM0522, the two reading frames were separated, and a stop codon in frame with the cryptic start site of STM0522 that remained in STM0521 was introduced. A new ribosome-binding site was also placed in between the two genes to allow for the translation of STM0522.

Original sequence (STM0521 in blue, STM0522 in green, overlap shaded in grey):

ATGCGTGAAAAGAGGGATTTATGGAACATCAAAGAGAGCTATACCAGCAGCGCGGTTATAGCGAAGACTTATTACCTAAGACAGAAACCCA
ACGAAACTGGAAGGCATTTAACTATTTCACCTTATGGATGGGATCTGTACATAACGTGCCAAATTACGTTATGGTCGGCGGTTTTTTTATACTG
GGGCTATCAACATTTAATATCATGTTGGCCATTATTATCAGCGCATTATTTATTGCGGCGGCGATGGTAATGAATGGCGCGGCAGGCAGCAAA
TATGGCGTTCCTTTTGCTATGATATTGCGAGGTTCTTACGGTGTCCGCGGCGCGCTATTCCCTGGATTATTGCGAGGGGGAATCGCGGCAATT
ATGTGGTTCGGCTTACAGTGTTACGCGGGATCGCTGGCATTTCTTATTTTGATTGGGAAGATCTGGCCAGGATTTCTGACATTAGGCGGAGAT
TTCAAGCTGCTGGGTCTTTCACTGCCAGGACTAATTACTTTTCTAATTTTTTGGATTATTAACGTTGGCATCGGTTTTGGCGGTGGTAAAGTATT
AAATAAATTTACCGCTATCCTCAATCCATGTATTTACATTGTCTTTGGCGGCATGGCTATTTGGGCAATATCGCTGGTCGGCATTGGCCCGATT
CTGGACTATCTGCCTTCAGGCGTGCAAAAAGCAGAGCACAGCGGCTTTCTGTTCCTGGTGGTGATTAACGCCGTAGTCGCCGTCTGGGCTGC
GCCAGCGGTAAGCGCGTCCGATTTCACGCAAAACGCGCATTCATTTCGCGCTCAGGCATTGGGACAAACATTGGGGCTTATCGTAGCGTATA

TATTATTCGCCGTAGCCAGCGTGTGCATTATTGCCGGAGCCAGTATTCATTATGGTATGGATACCTGGAACGTGCTGGATATTGTGCAGCGCT
GGGACAGCTTGTTTGCTTCATTCTTTGCGGTGCTGGTGATTCTGATGACGACAATTTCAACCAACGCCACCGGTAATATTATACCTGCGGGGT
ATCAAATTGCGGCGCTTGCCCCGACAAAGCTTAACTATAAAAATGGCGTAATGATTGCCAGTATTATCAGTCTACTGATTTGCCCATGGAAATT
AATGGAGAATCAGGACAGTATTTATCTGTTCCTCGATATTATCGGCGGTATGCTTGGCCCGGTAATCGGCGTAATGTTAGCTCATTATTTTGTG
GTAATGCGTGGAAAAATTAATCTTGATGAGCTGTACACCGCCAGCGGTGATTACAAATATTATGATAATGGATTTAACCTGACTGCATTTTCA
GTCACCCTAGTCGCAGTCATTCTGTCATTAGGCGGCAAGTTTATACCATTTATGGAGCCTTTATCCCGCGTGTCCTGGTTTGTAGGCGTAATAG
TTGCGTTCGTTGCTTATGCGCTATTGAAGAAACGCACGGGTTTTGAAAATACAGGAGAGAAAAAACTCGCAGGTTAA

New sequence (extra DNA sequence shaded in grey):

ATGCGTGAAAAGAGGGATCTATGGAACATCAAAGAGAGCTATACCAGCAGCGCGGTTATAGCGAAGACTTATTACCTAAGACAGAAACCCA
ACGAAACTGGAAGGCATCTAACTATTTCACCTTATGGATGGGATCTGTACATAACGTGCCAAATTACGCTATGGTCGGCGGTTTTTCTATACTG
GGGCTATCAACATCTAATATCATGCTGGCCACTACTATCAGCGCACTATCTACTGCGGCGGCGATGGTAACTTAAGGAGGGATCTGTACATAA
CGTGCCAAATTACGTTATGGTCGGCGGTTTTTTTATACTGGGGCTATCAACATTTAATATCATGCTGGCCATTATTATCAGCGCACTATTTATTG
CGGCGGCGATGGTAATGAATGGCGCGGCAGGCAGCAAATATGGCGTTCCTTTTGCTATGATACTGCGAGGTTCTTACGGTGTCCGCGGCGCG
CTATTCCCTGGACTACTGCGAGGGGGAATCGCGGCAATTATGTGGTTCGGCCTACAGTGTTACGCGGGATCGCTGGCATTTCTTATTCTGATT
GGGAAGATCTGGCCAGGATTTCTGACACTAGGCGGAGATTTCAAGCTGCTGGGTCTTTCACTGCCAGGACTAATTACTTTTCTAATTTTTTGGA
TTATTAACGTTGGCATCGGTTTTGGCGGTGGTAAAGTACTAAATAAATTTACCGCTATCCTCAATCCATGTATTTACATTGTCTTTGGCGGCAT
GGCTATTTGGGCAATATCGCTGGTCGGCATTGGCCCGATTCTGGACTATCTGCCTTCAGGCGTGCAAAAAGCAGAGCACAGCGGCTTTCTGTT
CCTGGTGGTGATTAACGCCGTAGTCGCCGTCTGGGCTGCGCCAGCGGTAAGCGCGTCCGATTTCACGCAAAACGCGCATTCATTTCGCGCTCA
GGCACTGGGACAAACACTGGGGCTTATCGTAGCGTATATACTATTCGCCGTAGCCAGCGTGTGCATTATTGCCGGAGCCAGTATTCATTATGG
TATGGATACCTGGAACGTGCTGGATATTGTGCAGCGCTGGGACAGCCTGTTTGCTTCATTCTTTGCGGTGCTGGTGATTCTGATGACGACAAT
TTCAACCAACGCCACCGGTAATATTATACCTGCGGGGTATCAAATTGCGGCGCTTGCCCCGACAAAGCTTAACTATAAAAATGGCGTAATGAT
TGCCAGTATTATCAGTCTACTGATTTGCCCATGGAAACTAATGGAGAATCAGGACAGTATTTATCTGTTCCTCGATATTATCGGCGGTATGCTT
GGCCCGGTAATCGGCGTAATGCTAGCTCATTATTTTGTGGTAATGCGTGGAAAAATTAATCTTGATGAGCTGTACACCGCCAGCGGTGATTAC
AAATATTATGATAATGGATTTAACCTGACTGCATTTTCAGTCACCCTAGTCGCAGTCATTCTGTCACTAGGCGGCAAGTTTATACCATTTATGG
AGCCTCTATCCCGCGTGTCCTGGTTTGTAGGCGTAATAGTTGCGTTCGTTGCTTATGCGCTACTGAAGAAACGCACGGGTTTTGAAAATACAG
GAGAGAAAAAACTCGCAGGTTAA

SI 2.4 Deleted regions of the *Salmonella* genome

The deletions made in the *Salmonella* genome (based on AE006468.1) to reduce the overall genome size and to remove unstable regions (eg. transposable elements), selected pathogenic regions (eg. SPI-1 and SPI-2), and a number of open reading frames annotated as pseudogenes. The overall genome size was reduced by 386761 bp, from 4857450 to 4470689 bp.

Table 2.4.1. Genes and multi-gene regions of the *Salmonella* genome that have been removed in the final recoded genome design. In the case of individual genes, only the open reading frame was removed. The intergenic regions were only removed in multi-gene regions as signified by the naming First-gene:Last-gene.

Deleted region	Coordinates in recoded genome
STM0241	281,688
STM0284	322,690
STM0297	339,617
STM0314	357,738
STM0436A:STM0326	366,931

mod	399,632
res	399,641
STM0479	529,299
STM0555	604,361
STM0556	604,395
STM0560	608,122
STM0767	819,860
hutI:hutG	839,797
hutC	839,841
STM0790	840,037
hutH	840,038
fels-1	944,474
STM0946	965,364
STM0947	965,826
gifsy-2	1,037,283
pipA	1,069,116
pipB	1,069,337
STM1089	1,069,554
pipC	1,069,563
sopB	1,069,579
orfX:STM1093	1,069,910
pipD	1,070,018
STM1120	1,094,077
STM1186	1,159,527
sifA	1,197,348
phoQ:phoP	1,203,719
STM1245	1,215,831
STM1248	1,217,707
orf319	1,358,117
orf242	1,358,289
ssrB	1,358,467

ssrA	1,358,497
ssaB	1,358,897
ssaC:ssaD	1,358,898
ssaE	1,358,905
STM1397	1,359,168
STM1398:sscA	1,359,174
sseC	1,359,176
sseD	1,359,191
sseE	1,359,193
sscB	1,359,244
sseF:sseG	1,359,259
ssaG	1,359,352
ssaH	1,359,392
ssaI:ssaJ	1,359,403
STM1410:ssaL	1,359,420
ssaM:ssaN	1,359,477
ssaO:ssaQ	1,359,479
ssaR:ssaU	1,359,546
STM1464	1,396,776
STM1465	1,396,785
STM1474	1,408,568
STM1553	1,486,171
sifB	1,548,754
STM1629-STM1638	1,575,236
STM1666	1,604,932
sopE2	1,797,356
STM1860	1,801,475
STM1862-STM1871	1,803,263
STM1930	1,861,002
STM1957	1,880,833
STM2003	1,918,131

STM2132	2,059,848
STM2210	2,144,194
STM2471	2,413,032
STM2509	2,455,534
sinH	2,481,494
gifsy-1	2,559,664
STM2664	2,590,273
STM2689	2,610,221
fels-2	2,615,594
STM2765	2,643,426
STM2768:STM2769	2,646,774
STM2778-STM2782	2,661,024
avrA	2,740,430
sprB	2,740,590
hilC	2,740,974
STM2868:prgJ	2,741,318
prgM	2,741,336
hilA	2,744,874
sptF	2,744,891
sicP	2,746,584
STM2880	2,746,610
iacP	2,746,653
sipA	2,746,671
sipD	2,746,689
sipC	2,746,759
sipB	2,746,786
sicA	2,746,788
spaS:spaW	2,746,925
spaP:invB	2,747,214
invA	2,747,237
invE:invF	2,747,261

STM2942	2,785,243
sopD	2,789,919
STM2975	2,828,688
STM3032	2,894,817
STM3302	3,171,447
STM3460	3,310,883
STM3479	3,333,999
STM3520	3,381,036
STM3530	3,392,207
STM3654	3,536,236
sugR:STM3754	3,646,626
rhuM	3,646,812
rmbA	3,648,110
misL	3,648,184
fidL	3,648,278
marT	3,648,288
STM3760	3,648,340
slsA	3,648,957
cigR	3,649,185
mgtB	3,649,499
mgtC	3,649,718
STM3766	3,651,171
STM3767:STM3768	3,651,406
STM3769:STM3770	3,651,409
STM3771	3,651,554
STM3772	3,651,564
STM3773	3,651,782
STM3774	3,651,978
STM3775	3,652,083
STM3779	3,654,308
STM3780	3,654,317

STM3781:STM3782	3,654,379
STM3783:STM3784	3,654,454
STM3785	3,654,471
STM3806	3,676,464
STM3844	3,718,833
STM3916	3,793,180
STM3945	3,819,016
STM4140	4,026,904
STM4195	4,083,789
STM4196-STM4218	4,084,050
STM4219	4,084,761
STM4257:STM4260	4,124,282
STM4261	4,124,298
STM4262	4,124,337
STM4311	4,182,493
STM4454	4,319,420
STM4457	4,323,712
STM4490	4,356,328
hsdS:hsdM	4,397,369
hsdR	4,397,526

SI 2.5 Mutation of LguI restriction sites

All 754 instances of the LguI restriction site (5'-GCTCTTC-3') were mutated by searching all possible single base permutations for synonymous mutations when arising within coding regions. For sites in intergenic regions, a BLAST comparison of homologous regions in related organisms was used to guide the design: naturally occurring SNPs served as exemplary alternatives from which a candidate mutation was picked.

In one instance, an LguI site was present in two overlapping reading frames. Similar to when recoding codons, PROVEAN was used to predict a suitable mutation with minimal impact. In this case, an E332G mutation was introduced in STM3276.

A full list of LguI changes is available in SI File 2 (Tab 2: LguI changes).

SI 2.6 Analysis for putative list of essential genes

To ensure that resistance markers did not disrupt essential genes during strain construction, a list of putative essential genes was generated. Since no such list of essential genes has been published for *S. typhimurium* LT2, a comprehensive list was created using various sources of information. Genes essential to other *Salmonella* strains (ATCC 14028, SL3261 and SL1344) were obtained from the literature(Barquist et al., 2013; Knuth, Niesalla, Hueck, & Fuchs, 2004; Rolfe et al., 2012). This list was supplemented with genes essential to *Escherichia coli* obtained from the Keio collection(Baba et al., 2006) and the *E. coli* wiki (http://ecoliwiki.net). Based on their annotation, the corresponding genes in *S. typhimurium* LT2 were identified, yielding a total of 436 genes. Additionally, all nucleotide sequences of these genes were obtained from their respective GenBank files, and subsequently translated into protein sequences that were used to perform a protein BLAST against all translated coding sequences of *S. typhimurium* LT2 (Figure 6). For strain ATCC 14028, no genome sequence could be obtained, and the sequences had to be manually obtained from the genomic sequence of the closely related strain ATCC 14028S, since documentation on the exact difference between the two strains could not be found in literature. A lenient cut-off value of E = 1e-90 was used to minimize the risk of missing essential sequences. For all essential sequences for which no homolog was found using this approach, as well as for all essential non-protein coding sequences, nucleotide sequences were obtained and a nucleotide BLAST was performed. Finally, all tRNA- and rRNA-encoding regions as annotated in the *S. typhimurium* LT2 GenBank file were appended to the list, yielding a total of 613 potentially essential sequences. The list of essential sequences can be found in SI file 2 (Tab 3: putative essential genes).

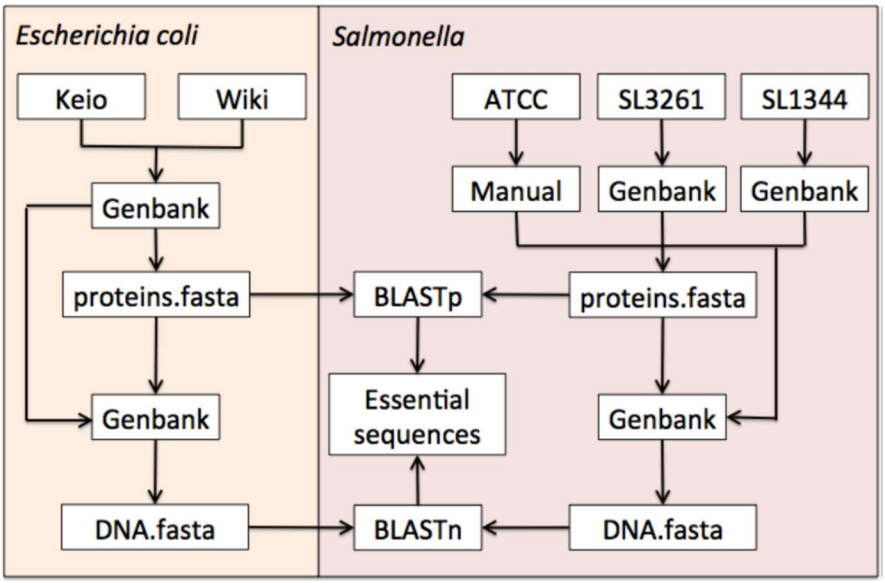

Figure 2.6.1 Schematic representation of the pipeline used to generate a list of putative essential genes in *S. typhimurium* LT2. Essential genes in *E. coli* were obtained from the Keio collection(Baba et al., 2006) and the E. coli wiki, those for various *Salmonella* strains were obtained from the literature(Barquist et al., 2013; Knuth et al., 2004; Rolfe et al., 2012). Essential nucleotide sequences were obtained through their respective GenBank files, which were translated to protein sequences. Both BLASTp and BLASTn were used to obtain putative essential genes in *S. typhimurium* LT2.

SI 3. Details of DNA constructs

SI 3.1 Design of recoded constructs for SIRCAS

Regions A and B of the recoded genome were split into 10-25 kb segments, each overlapping with the adjacent segment by 1 kb to provide homology for the swapping of selection markers by SIRCAS. Based on the list of putative essential genes, each segment was designed to finish at a location in the genome where insertion of a selection marker would be viable (not within or 250 bp upstream of a putative essential gene).

Each segment was then divided into pieces with a maximum size of 4 kb to comply with synthesis constraints. Adjacent pieces were given 100 bp overlaps for assembly in yeast, taking into account that some of this homology would be lost in downstream processing.

Gen9 DNA was clonal and sequence verified, whilst DNA from SGI-DNA/BioXp was not clonal nor sequence verified.

DNA designs were first submitted to Gen9 for synthesis. As linear DNA produced by Gen9 was flanked by mandatory 36 bp sequences, recognition sequences for type IIs restriction enzymes were placed at both ends of each piece to enable removal of the flanking sequences. These restriction enzymes were chosen such that their recognition sequence was not present within the piece itself. After restriction digest of the flanks and subsequent heat inactivation of the enzyme, the DNA was used directly for assembly in yeast without further purification. In cases where heat inactivation was not possible, the DNA was purified using a DNA Clean and Concentrator kit (Zymo Research) prior to assembly.

In addition, there were several forbidden restriction sites for the Gen9 manufacturing process: BsaI and AarI. For each construct containing more than the allowed number of these restriction sites, the offending restriction site was manually edited to change the sequence. When the site was present within a coding region, a single base was edited to create a synonymous codon change. When more than one synonymous codon was available, the codon with the closest frequency of occurrence in the *Salmonella* genome was chosen. When changes were made in intergenic regions, homologous regions in other related organisms were searched by BLAST to determine if any changes were naturally occurring in these organisms. These were implemented where possible, and where no suitable mutations were found by BLAST, an arbitrary base change was made. These manual changes are listed in SI file 2 (Tab 6: manual changes).

Designs that could not be synthesized by Gen9 were ordered from SGI-DNA as reagents for the BioXp 3200 machine. To comply with synthesis constraints, the previously designed <4kb fragments were typically split in half to reduce the size to a maximum of 2 kb, whilst maintaining 100 bp overlaps. DNA from the BioXp 3200 were run on a 1% agarose gel to check for the correct band size. In cases where byproduct bands were observed, the DNA was purified by gel extraction using a Zymoclean Gel DNA Recovery Kit (Zymo Research) prior to assembly.

Repeat regions such as tRNAs that could not be synthesized were amplified from the genome by PCR if not recoded (ie. not encoding for a protein), or synonymous mutations were introduced to change the repetitive sequence when in coding regions. A BLAST approach was utilized for repeat regions in intergenic regions.

Sequences of all synthetic constructs are available in SI file 3 (recoded DNA constructs.fasta). Full details of the DNA assemblies, restriction enzymes involved, commercial sources, integration loci in the recoded genome are listed in SI file 2 (Tab 1: recoded DNA fragments).

SI 3.2 PCR conditions and primers used for DNA construction

Table S3.2.1. Primers used for the construction of DNA for the integrated recombineering element, the YAC backbone for DNA assembly, and the counterselectable marker.

Primer name	Template	Sequence
IRE-Gent-F	IDT g-block (synthetic DNA)	cctctggtaaggttgggaagcccgggccctactcaaagtaatccatatcgagctttg gcaacttgagtaaacgtgaggatacgccaatatcattatcaatcatcagatgcacc agcctgtcgccatccTTAGGTGGCGGTACTTGGGTCGATATC
IRE-Gent-R	IDT g-block (synthetic DNA)	atgcttcaataatattgaaaaaggaagagtAACTCCCATGTGGGAGTTTT TTTTTGACTC
IRE-pKD46-F	pKD46	ACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGG
IRE-pKD46-R	pKD46	caggctttttacgcgaggttttttaccgccgctggcggcgcgttcggccttaattttt tccagcagggcggcggcgctgttttcaccgctgataagtTCATCGCCATTGCT CCCCAAATAC
IRE-F	overlap PCR product	CTACTCAAAGTAATCCATATCGAGCTTTGGCAAC
IRE-R	overlap PCR product	CAGGCTTTTTTACGCGAGGTTTTTTTACCG
YAC-F	pYAC4 plasmid	gaagggatgctaaggtagagggtgaacgttacagaaaagcaggctgggaagcat atttgagaagatgcggccagcaaaacAGGCGTATCACGAGGCCCTTTC
YAC-R	pYAC4 plasmid	ACCGGCATAACCAAGCCTATG
relE-F	pSLC-242	cttgcctcagggatacgatcgtatccctgaatatttcccaAGGAACTTCGGAA TAGGAAC
relE-R	pSLC-242	aaatatattagtgagaatgattttactcctgagttttgaAACGGAGTAGAGAC GAAAGTG

All standard PCR was carried out using the Q5 High-Fidelity 2X Master Mix (New England Biolabs) according to manufacturer's protocols.

Multiplex PCR was carried out using a Multiplex PCR Kit (Qiagen) in a 10 µL volume, using 0.125 µL of each primer in a 10 µM stock. Primer pairs were designed to generate PCR products between 100 bp and 1.5 kb. Typically, multiplex PCR directly from yeast colonies containing DNA assemblies were run using the following program: 15 min at 95 ^{o}C; 30 cycles of 30 s at 95 ^{o}C, 90 s at 54 ^{o}C, 60 s at 68 ^{o}C; 5 min at 68 ^{o}C.

SI 3.3 Construction of the integrated recombineering element

The lambda recombineering functions were amplified from the plasmid pKD46 using primers IRE-pKD46-F and IRE-pKD46-R. A gentamicin resistance cassette was amplified from a synthetic DNA using primers IRE-Gent-F and IRE-Gent-R. Overlap-extension PCR was then conducted using the two previous PCR products and primers IRE-F and IRE-R to give the complete recombineering element ready for genomic integration.

The recombineering element was integrated into the *Salmonella* LT2 strain TT22971, kindly provided by the laboratory of Prof. John Roth (UC Davis). The genotype of this strain is: metA22 metE551 trpD2 ilv-452 leu- pro-(leaky) hsdLT6 hsdSA29 hsdB strA120 $pKD46 araC bla oriR101 repA101ts lambda red (gam+ bet+ exo+)

SI 3.4 Construction of the YAC backbone for DNA assembly

The vector backbone for DNA assembly in yeast was obtained as a PCR product from the plasmid pYAC4(Burke, Carle, & Olson, 1987) using primers YAC-F and YAC-R. The resulting 2720 bp construct contains the ARS1/CEN4 elements for replication and a TRP1 selection marker.

SI 3.5 Rhamnose-inducible counterselection constructs

Plasmids pSLC-217 and pSLC-242 were a gift from Swaine Chen (Addgene plasmid #73163 and #73189)(Khetrapal et al., 2015), containing KanR-relE and CamR-relE cassettes respectively.

As an alternative to screening with conventional positive selection markers, we also demonstrated a counterselection strategy to select for the desired phenotype. Utilizing a counterselectable marker based on rhamnose induction of the toxin relE(Khetrapal et al., 2015), the correct phenotype could be selected for by plating strains on antibiotic media supplemented with rhamnose.

The counterselection strategy was used on recoded strain A7 to swap the Kan resistance cassette with a CamR-relE cassette. The CamR-relE integration cassette for A7 was generated from pSLC-242 by PCR using primers relE-F and relE-R. Integration of this cassette by recombineering gave an A7 strain containing CamR-relE, which was found to have significantly impaired growth on LB plates containing 2% rhamnose. Performing the next round of SIRCAS to obtain the A8 recoded strain was successful, and recoded colonies which lost the previous CamR-relE marker could be identified as colonies with a regular growth phenotype.

We note that the original study performed counterselection on minimal medium, but we found that whilst not lethal, the relE toxin was able to suppress growth sufficiently to allow for easy identification of relE-negative colonies.

SI 3.6 Characterization of DNA assembled in yeast

The following data is an example of multiplex PCR characterization of DNA assemblies as performed on assembly A2.

Table S3.6.1. Primers for multiplex PCR analysis of assembly A2.

Primer name	Sequence
MP-21	ACTCCAAGCTGCCTTTGTGTG
MP-22	AGTTCGTGACTGGGGTGAGC
MP-23a	AGCTTTAATGCGGTAGTTTATCACAG
MP-23	ACACCTGATAGAAGAGTTCGTGG

MP-24	ATGCCAGTAACAATGCCGAAG
MP-25	CAGGTATTGGTTGCGTCAGC
MP-26	TGTGATTCGAGGTCAGGTAGC
MP-27	TTCAGCACCGTGAAGCAG
MP-28	TGACAGCCAGCCCATTCATTC
MP-29	TTATCCCGTTGATTACCAGACAC

Table S3.6.2. Primer pairs and product sizes for multiplex PCR analysis of assembly A2. Each pair amplifies a region that covers at least one assembly junction.

Primer pair	PCR product size (bp)
MP-21/22	1154
MP-28/29	793
MP-24/25	423
MP-26/27	281
MP-23/23a	207

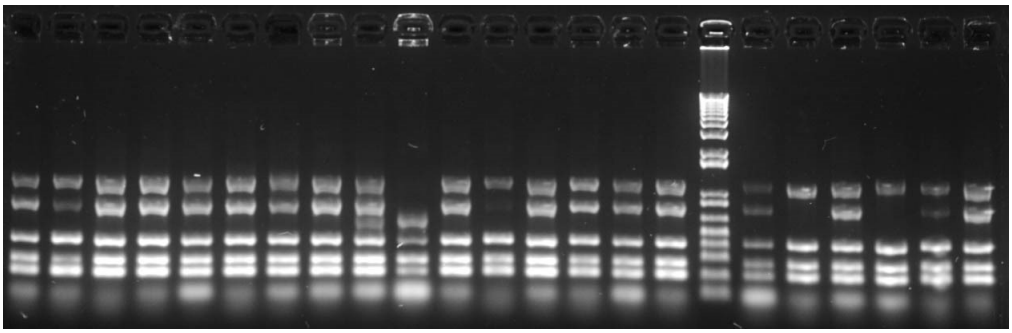

Figure S3.6.1. Gel of multiplex PCR analysis of assembly A2. Lane 17 is the 1 kb plus ladder, whilst the remaining 22 lanes are different yeast colonies obtained from the DNA assembly protocol. In this case, 17/22 of the colonies have the correct set of five PCR product bands.

The following data shows examples of diagnostic digests of DNA assemblies amplified by RCA.

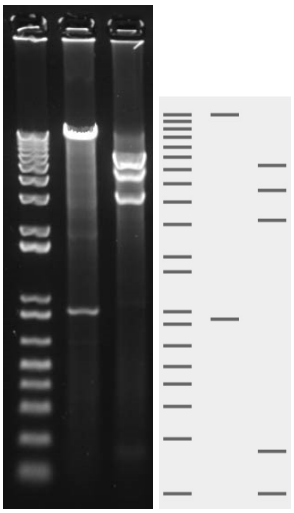

Figure S3.6.2. Left: Gel of diagnostic digests of assembly A3, using enzymes EcoRI and BsrGI in Lanes 2 and 3. The 1 kb plus ladder is shown in Lane 1. Right: Theoretical expected bands after diagnostic digest.

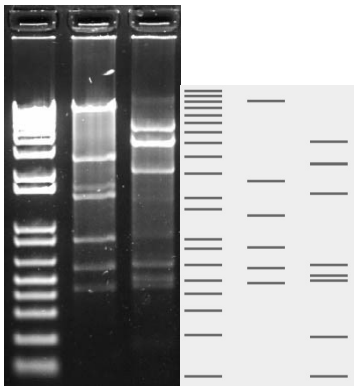

Figure S3.6.3. Left: Gel of diagnostic digests of assembly A4, using enzymes EcoRV and BpiI in Lanes 2 and 3. The 1 kb plus ladder is shown in Lane 1. Right: Theoretical expected bands after diagnostic digest.

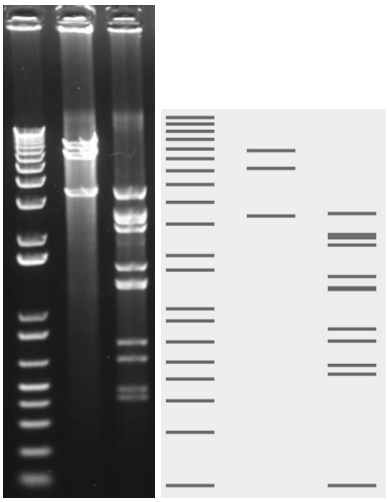

Figure S3.6.4. Left: Gel of diagnostic digests of assembly A5, using enzymes BpiI and EcoRV in Lanes 2 and 3. The 1 kb plus ladder is shown in Lane 1. Right: Theoretical expected bands after diagnostic digest.

Uncropped versions of the previous gels are shown here:

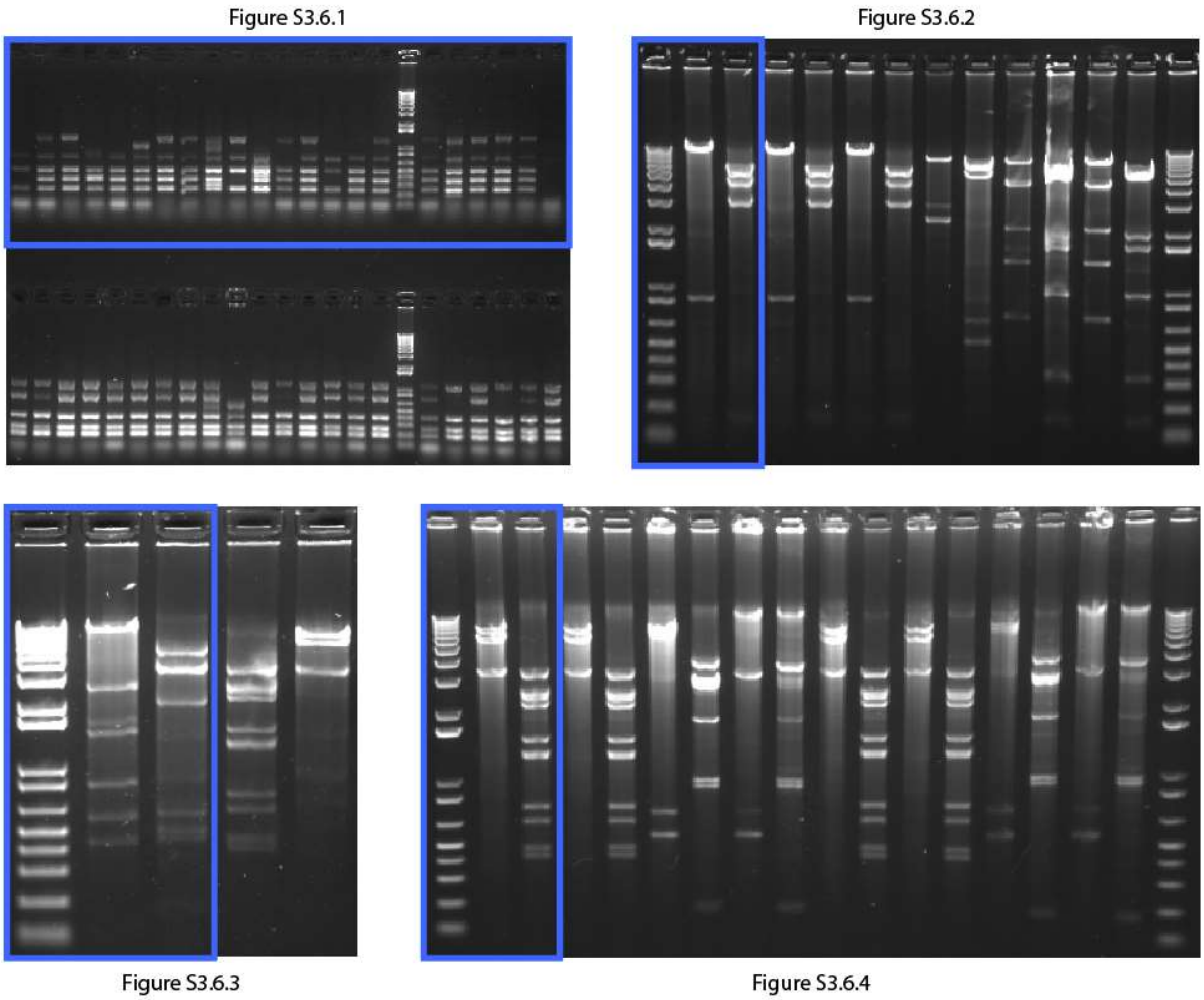

Figure S3.6.5. Uncropped versions of the previous four gel figures are shown here. The area used in the final figure is shown in the blue box.

SI 4. Further details of the SIRCAS protocol

SI 4.1 Timeline for SIRCAS workflow

Each round of the SIRCAS protocol can be comfortably completed in two days, according to the following schedule.

Table S4.1.1. Schedule for completing one round of SIRCAS in two days.

Timing	Activity

DAY 1	
0 h	Inoculate cultures for recombineering
2.5 h	Prepare competent cells
3.5 h	Electroporate cells
4 h	Outgrowth of cells
9 h	Plate cells on selective media overnight
DAY 2	
24 h	Streak out colonies on double antibiotic plates to identify those with correct phenotype. At the same time, restreak onto single antibiotic plates to obtain single colonies.
30 h	At this stage, correct colonies are identifiable but single colonies are difficult to pick. PCR an internal segment and submit for overnight Sanger sequencing.
33 h	Inoculate several different single colonies with correct phenotype in overnight cultures for next round of SIRCAS (choosing which to continue with the next morning based on sequencing results).

We have integrated up to 25 kb segments of linear DNA (corresponding to 23 kb recoding) in one step. Therefore, it is possible to integrate at least 75 kb of DNA into each strain per week, and potentially more if the process is automated. The process is also easily parallelized to integrate DNA into multiple strains with each round of SIRCAS.

SI 4.2 Data for percentage of correct phenotype after SIRCAS

Table S4.2.1. Percentage of colonies with the correct marker phenotype after each round of SIRCAS. This data corresponds to the graph in Figure 3c in the main text.

Segment integrated	Correct phenotype	Total screened	Percentage correct
A1	12	140	8.6
A2	5	60	8.3
A3	10	60	16.7
A4	8	50	16.0
A5	6	48	12.5
A6	19	68	27.9
A7	15	64	23.4
A8	9	200	4.5
A9	3	32	9.4
A10	4	47	8.5

A11	17	48	35.4
A12	31	200	15.5
A13	23	141	16.3
B1	1	28	3.6
B2	18	44	40.9
B3	1	31	3.2
Average	-	-	**15.7**

SI 4.3 Sanger sequencing of internal regions for each round of SIRCAS

Although the correct phenotype is indicative of complete recoding, Sanger sequencing an internal portion of each recoded segment was performed to check for the possibility of multiple internal crossover events.

Sanger sequencing was performed by Eton Bioscience or Genewiz.

Table S4.3.1. Number of colonies which were Sanger sequenced internally after SIRCAS to check for the possibility of multiple internal crossover events. In the cases were only one colony was sequenced, this data was obtained from whole genome sequencing rather than targeted Sanger sequencing.

Segment	Correctly recoded	Total sequenced	Percentage recoded
A1	1	1	100
A2	1	1	100
A3	3	2	66.66667
A4	15	13	86.66667
A5	3	3	100
A6	19	17	89.47368
A7	4	3	75
A8	9	4	44.44444
A9	3	3	100
A10	4	4	100
A11	17	12	70.58824

A12	8	7	87.5
A13	10	10	100
B1	1	1	100
B2	1	1	100
B3	1	1	100
Total	**100**	**83**	-

Table S4.3.1. Primers used for Sanger sequencing internal regions for each round of SIRCAS. The same primers were used for amplification from the genome and subsequent sequencing.

Primer name	Sequence
A3-F	AGGTTCTGTTTATGTCTGCCTGTG
A3-R	CCCAGTTCGCCCAGTATATCTCAC
A4-F	GCAGTATTTTGACGACGAGACAGG
A4-R	ATGGGTTATGCTATCTTCCTGACCTC
A5-F	GGTGAAAGTCATCCGTTCC
A5-R	CAGATTGCCCAGATGCCTC
A6-F	CGATGCTAACGCTTGATATGCC
A6-R	AATCGGCACTACGGTCCAGTTG
A7-F	ATGTTCGGTCACTACTTCAGGCAG
A7-R	TCATTGGTATTTTCAGGCTTCAGCAG
A8-F	TTTAATGACTATGTGAGCCTGTACG
A8-R	AGCGTTTGAAGCCATATTTACGAC
A9-F	AATGTTATTCAGACTGCTCATGAC
A9-R	CGCTGGTTGTTATTGCC
A10-F	ACCGTCTTCTTTCCAGGTATCG
A10-R	TGGAAGTAATCGAGCCATGAATAACTG
A11-F	TGACGTATCGCACCTCTGGAAC
A11-R	CGTTGCCCAGGTACATAAATGACTG
A12-F	CCAGCCAGACCAGATTATCCAG
A12-R	CCCTAACGCAACAACTACAGC
A13-F	CCGAAGGTGACGTAGTGGTATTC
A13-R	TTCCACTTCAGGCGTTTGTCG

SI 5. Details of the conjugative assembly protocol

SI 5.1 Donor strain construction

To construct the A13 donor strain, the first step was to swap the gentamicin resistance cassette used for the integrated recombineering element with a tetracycline resistance cassette, as gentamicin resistance is needed downstream to maintain the pTA-Mob plasmid. The tetracycline resistance cassette with

homology arms was generated using primers CP1-F/R. The template used was a gift from the Roth lab (*Salmonella* LT2 strain TT25401: $CRR2061[zfa-9223::kan*zfa-9228::tetRA*Peut] eut-38::MudA).

The origin of transfer was then integrated upstream of recoded region A. An origin of transfer was amplified from pRK24 plasmid(Guiney & Yakobson, 1983; Natalie J Ma, Moonan, & Isaacs, 2014) using primers CP2-F/R, whilst a spectinomycin resistance cassette was amplified from a pCDF plasmid using primers CP3-F/R. These two fragments were then joined by overlap PCR, and then integrated by recombineering into the *Salmonella* genome.

Finally, the pTA-Mob plasmid(Strand et al., 2014) was transformed into the strain to complete the donor construction.

Table S5.1.1. Primers used for constructing strains for conjugation.

Primer name	Template	Sequence
CP1-F	Roth lab strain TT25401	attgaaaaaggaagagtaactcccatgtgggagttttttttCATTAATTCCTAATTTTTGT
CP1-R	Roth lab strain TT25401	atcattatcaatcatcagatgcaccagcctgtcgccatccCTAAGCACTTGTCTCCTGTT
CP2-F	pRK24	GGCGAGATCACCAAGGTAGTCGGCAAATAATCATGGCTCTGCCCTCGGGCGGACCACGCC
CP2-R	pRK24	agctcccgcaggagctgtgaaagaggggtaaaggtacaaaCGTCTCTCGCCTGTCCCCTC
CP3-F	pCDF plasmid	cggaaggactggctcctgcctgggattaattacccgatccTTTGTTTATTTTTCTAAATA
CP3-R	pCDF plasmid	GGCGTGGTCCGCCCGAGGGCAGAGCCATGATTATTTGCCGACTACCTTGGTGATCTCGCC
CP4-F	pKD46	AGACCGTTGGCGCCACCCGGCAATATCGCAAACCGGATGGCAGGTGGCACTTTTCGGGGA
CP4-R	pKD46	CATATGTATATCTCCTTCTTAAATCTAGAGGATCTCTCGATTACCAATGCTTAATCAGTG
CP5-F	pUA66	GATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAATCGAGAGATCCTCTAGATTT
CP5-R	pUA66	TCCCACGCAGGAAGATAGTGGCTACCAGGCGGAAGCTTAGCTACTCAGGAGAGCGTTCAC

SI 5.2 Recipient strain construction

An ampicillin resistance + GFP construct was generated for insertion into the recipient B3 strain. The GFP cassette was included to assist with identifying gain and loss of the cassette. The ampicillin cassette was amplified from plasmid pKD46 using primers CP4-F/R. GFP was amplified from plasmid pUA66(Zaslaver et al., 2006) using primers CP5-F/R. The final construct was then generated by overlap PCR, then integrated into the Salmonella genome.

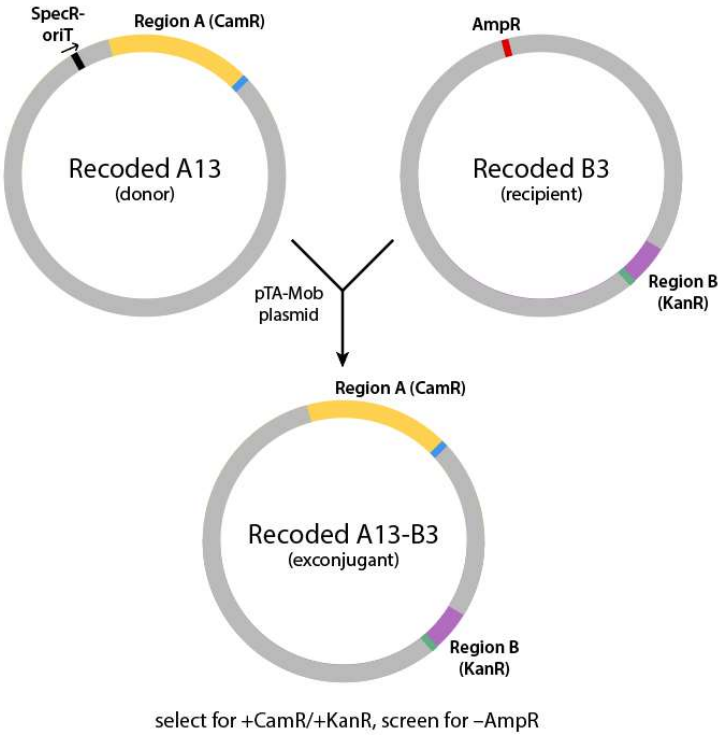

Figure S5.2.1. Recoded regions were hierarchically assembled using a conjugative approach. The 154 kb of recoding in strain A13 was successfully transferred into strain B3, using an integrated origin of transfer (oriT) upstream of region A, and the conjugation machinery on plasmid pTA-Mob(Strand et al., 2014). Successful exconjugants were identified by the presence of chloramphenicol and kanamycin resistance, and loss of ampicillin resistance, in a similar marker swapping approach used for the SIRCAS process.

SI 6. Sequence analysis of recoded *Salmonella* strain A13/B3

SI 6.1 Designed deletions made in recoded regions

Table S6.1.1. Deletions made in the genome, in addition to the 200 kb recoded. These designed deletions correspond to unwanted proteins in the final *Salmonella* design.

Deleted gene/region	Coordinates in recoded genome	Size of deletion (bp)	Annotation
STM0241	281,688	706	cutF pseudogene
STM0284	322,690	432	shiga-like toxin A subunit
STM0297	339,617	198	putative transposase
STM0314	357,738	1113	hypothetical protein
STM0436A:STM0326	366,931	1250	IS3 transposase, pseudogene
mod	399,632	1959	type-III restriction system
res	399,641	2976	type-III restriction system

SI 6.2 Alignment of Miseq sequencing reads to reference genome

All sequencing data was processed using Geneious (version 9.1.5)(Kearse et al., 2012). Reads were trimmed using an error probability limit of 0.05, paired, then aligned to either the reference wild-type *Salmonella* genome or the expected A13-B3 *Salmonella* reference genome, choosing to identify structural variants and deletions of any size, and using default sensitivity parameters (fine tuning: iterate up to 5 times, minimum mapping quality: 30, trim paired read overhangs, only map paired reads which map nearby, minimum support for structural variant discovery: 4 reads, maximum gaps per read: 10 %, maximum gap size: 15, word length: 18, index word length 13, maximum mismatches per read: 20%, maximum ambiguity: 4). The average coverage depth over the recoded regions of A13-B3 was 27 ± 9 reads, where the error is the standard deviation. SNPs/variations were identified using the default parmeters in Geneious 9.1.5, discarding any variations with p-values below 10^{-15}.

SI 6.3 Deviations from the recoded genome design in recoded regions

There are a number of different types of deviations from the recoded genome design (SI 3 – recoded Salmonella genome.gb). A large number of observed differences are due to manual changes made to satisfy DNA synthesis requirements (removal of restriction sites for BsaI and AarI, alterations to repeats, regions of extreme GC content). There are also several changes due to discrepancies between different versions of GenBank annotations (NC_003197.1 vs AE006468.1).

The remaining differences are errors arising from the genome recoding process – the vast majority of these originate from errors in non-clonal synthetic DNA, rather than the assembly and SIRCAS recoding process itself.

Out of 156 kb of the recoded regions that correspond to clonal sequence-verified synthetic DNA, only eight mutations are observed, which corresponds to an approximate error rate of 1/20000. Six mutations are single base-pair changes, whilst the remaining two are single base-pair deletions or insertions in tandem repeats. The yeast assemblies containing the regions corresponding to these eight mutations were Sanger sequenced and found not to contain the errors. Therefore, these errors must have arisen during RCA, integration or through spontaneous mutation.

Out of 44 kb of the recoded regions that correspond to non-clonal synthetic DNA, we observe 51 errors, which corresponds to an approximate error rate of 1/1000. There are 39 deletions, 1 single base-pair insertion at a tandem repeat, and 11 single base-pair changes.

A full list of sequence deviations and mutated proteins can be found at SI file 2 (Tabs 7/8: errors in recoded regions, proteins mutated by errors).

Minor morphological defects were observable in strains carrying the B2 segment, with a slight shortening of average cell size as seen by phase-contrast microscopy (Figures 7.2.1-19). Whilst recoding may be a contributing factor, we postulate that the phenotypic changes are most likely due to errors from non-clonal DNA used in the *rfb* region involved in O-antigen biosynthesis. Resulting changes in the bacterial cell surface may therefore be responsible for the observed phenotype.

Mutations in the rfb region (also known as the *wba* region) which may be relevant to the altered phenotype arising from segment B2 are shown in Figure S6.3.1. Single point deletions in B2 arising from

errors in non-clonal DNA introduce frameshift mutations in rfbM, rfbU, rfbV and rfbX, potentially impairing mannose activation, mannose and abequose attachment, and O-antigen translocation to the outer surface of the cell respectively (SI 1 Figure S6.3.1)(Reeves, Cunneen, Liu, & Wang, 2013). Altered O-antigens have previously been shown to accumulate within the cell and suppress growth as they are less efficiently translocated by rfbX (also known as *wzx*)(M. A. Liu, Stent, Hong, & Reeves, 2015), while it is also known that transposon-mediated disruption of several *rfb* genes (including rfbV and rfbX) is lethal in *S. typhimurium* strain SL3261(Barquist et al., 2013).

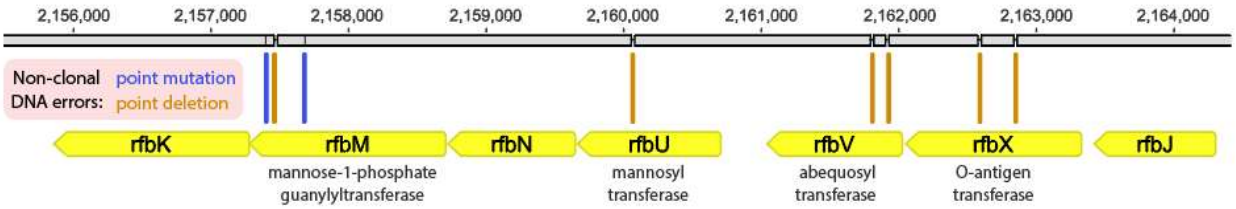

Figure S6.3.1. The differences in phenotype from segment B2 may be due to errors in four genes of the *rfb* region (O-antigen biosynthesis) arising from non-clonal DNA used during construction.

SI 6.4 Mutations in non-recoded regions

To investigate the possibility of other mutations arising in non-recoded regions of the genome, we conducted whole-genome sequencing of the integrated recombineering strain and recoded strains A13-B3, B3, A5 and A9. Consensus sequences were generated from the mapped reads using a strict cut-off of at least 75% of reads corresponding to the consensus. The reference sequence was used in any cases where less than three consistent reads were observed, and any cases with less than five cases. Whole genome alignments were carried out using the Mauve plug-in for Geneious(Darling, Mau, & Perna, 2010). The progressiveMauve algorithm was used with default settings (automatically calculate seed weight, and minimum LCB score, compute LCBs, full alignment) to find sequence discrepancies by genome-wide alignment. All discrepancies were manually examined and compiled. Discrepancies were discarded when arising from a call where the reference sequence was used due to low coverage, but the existing reads indicate a difference from the reference sequence. Ambiguous base calls, and ambiguous or low coverage regions near and within transposable elements were also not regarded as discrepancies.

A total of seven mutations from wild-type were found in all sequenced genomes. Two were found in both the strain recoded from B1-B3 and the final strain A13-B3. Although five more mutations were found in the intermediate strain recoded from segment A1 up to A9, these mutations accumulated during the recoding of the A region were not transferred into the A13-B3 strain during conjugation.

The list of these mutations can be found at SI file 2 (Tab 9: spontaneous mutations).

SI 7. Growth behavior of recoded *Salmonella* strains

SI 7.1 Growth curves

Full growth curves for recoded *Salmonella* strains at 37 °C in LB are shown in Figure 7.1.1. We note that strains B2, B3 and A13-B3 do not reach the same apparent maximum cell densities as the other strains. This effect is due to sedimentation effects in dense cultures during the plate reader assay (Figure 7.1.2).

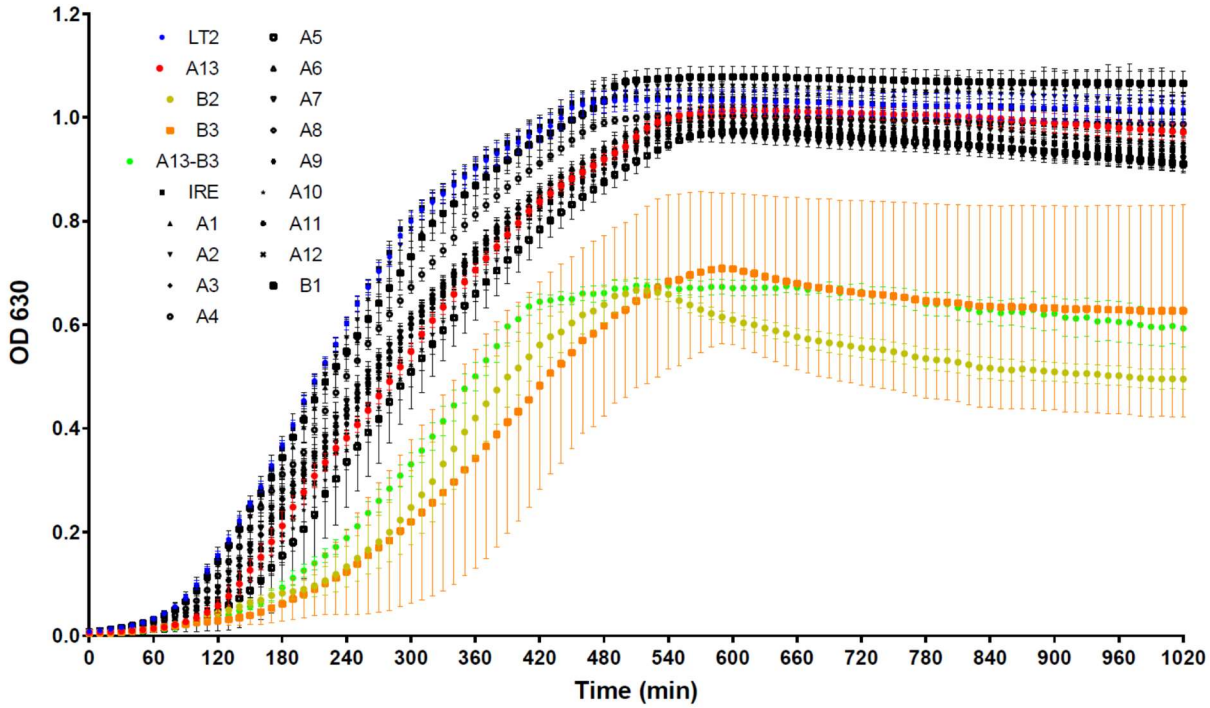

Figure S7.1.1. Growth curves for recoded *Salmonella* strains at 37 °C in LB, as measured on a plate reader. Each data point is the average of at least three biological replicates, and the error bars represent the standard deviation.

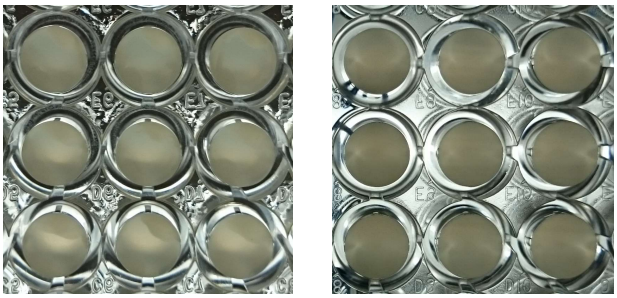

Figure S7.1.2. Left: Further examples of uneven sedimentation is observed at high cell densities under plate reader growth conditions, which may account for the larger variation and lower OD measurements for strains B2, B3 and A13-B3. Right: No uneven sedimentation is observed for cells not containing segment B2.

Similar growth trends were observed at 30, 37 and 42 °C. Cell growth was significantly impaired at 47 °C, and no trends were discernible. We note that the values calculated for B2, B3 and A13-B3 are compromised by uneven sedimentation under plate reader growth conditions.

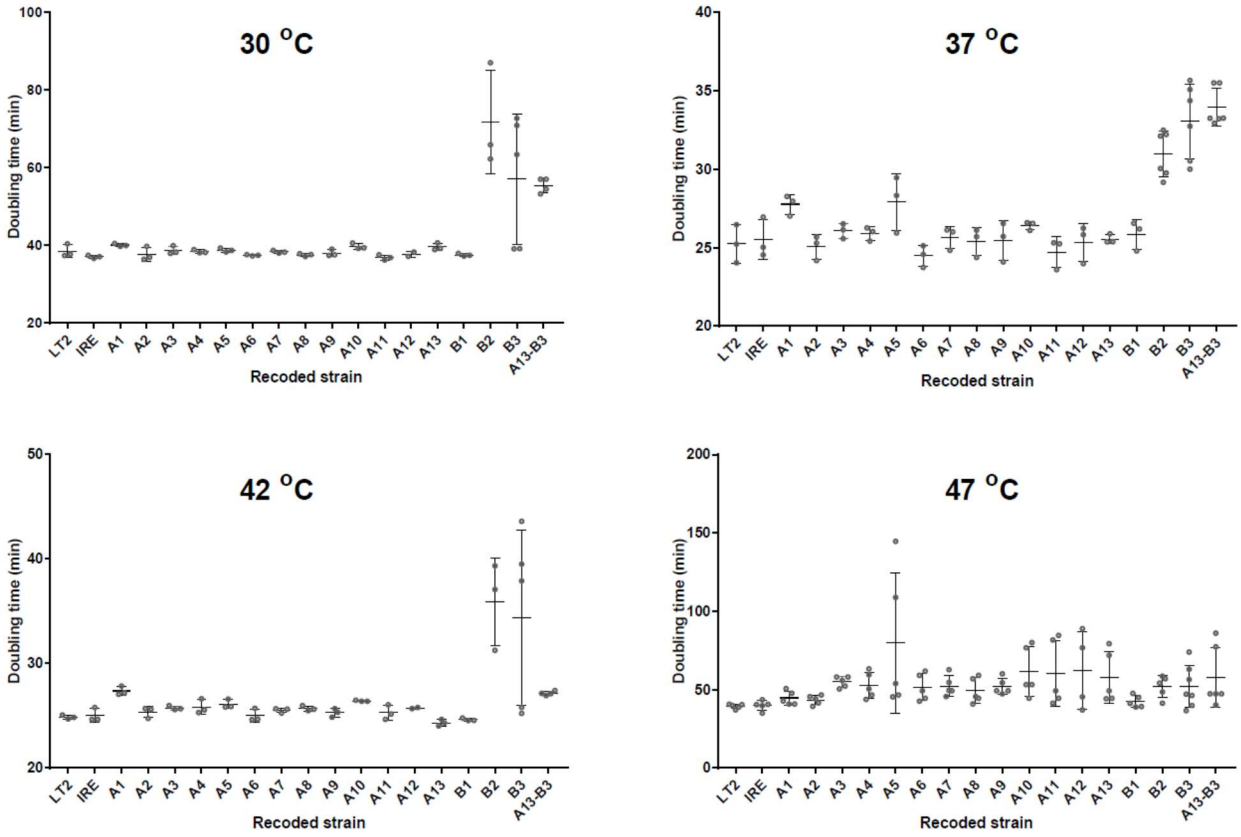

Figure S7.1.3. Doubling times for recoded strains as measured on a plate reader at different temperatures. Each data point is a different biological replicate, and the error bars represent the standard deviation.

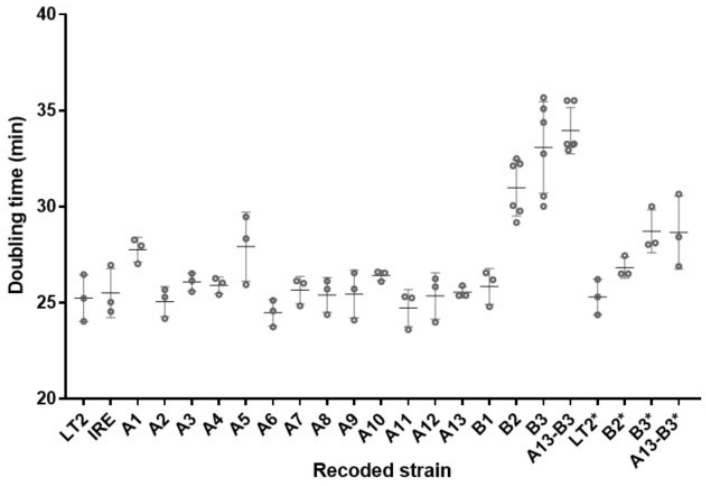

Figure S7.1.4. Doubling times for recoded strains, showing the flask-cultured strains alongside the plate reader results. Strains where data was obtained from flask growth are marked with an asterisk.

SI 7.2 Light microscopy

Microscopy images were captured on a Nikon TE2000 microscope using a 100x phase objective, Perfect Focus, and an Orca ER camera. Cells were grown overnight in 200 µL LB cultures in 96-well plates at 37 ^{o}C whilst shaking on a Titramax 1000 at 900 rpm with a 1.5 mm radius (Heidolph). Cells in stationary phase were imaged on 35 mm glass bottom uncoated dishes (MatTek) under an agar pad. Images were processed using Fiji.

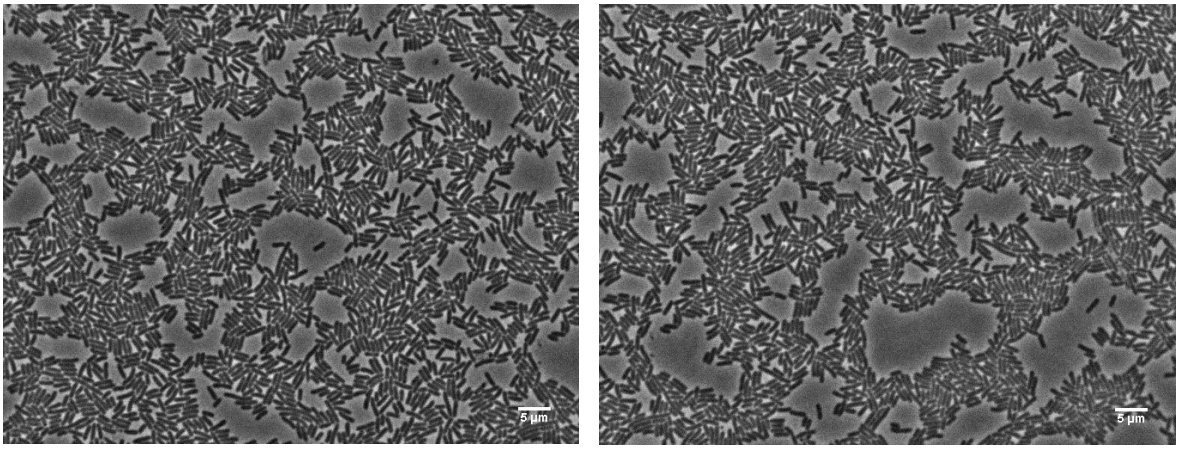

Figure S7.2.1. Light microscopy images of wild-type *S. typhimurium* LT2.

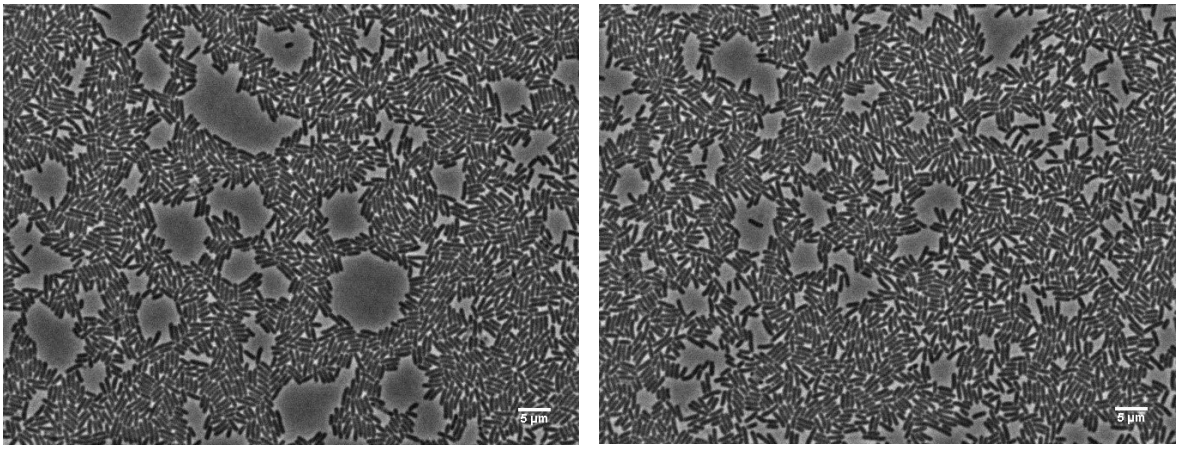

Figure S7.2.2. Light microscopy images of the integrated recombineering strain.

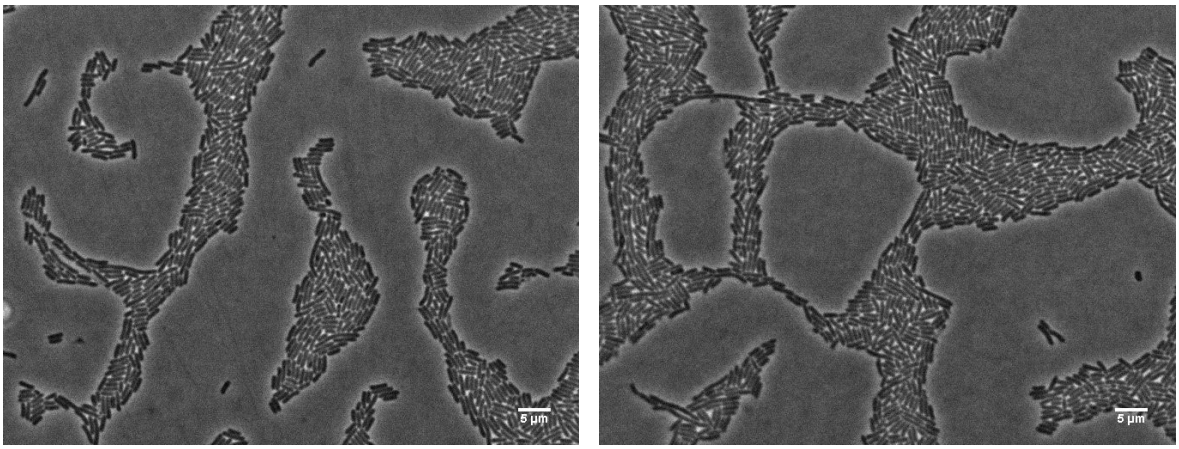

Figure S7.2.3. Light microscopy images of recoded strain A1.

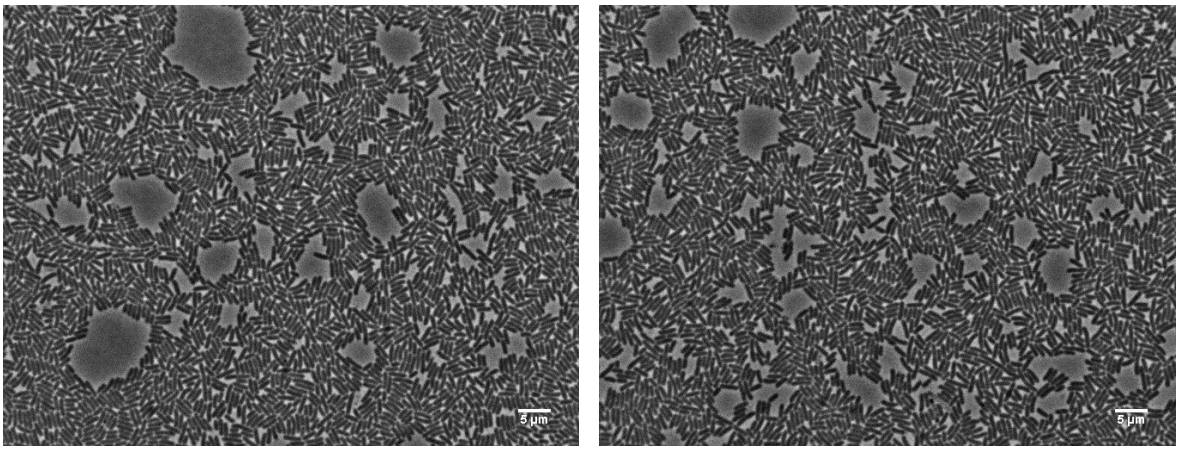

Figure S7.2.4. Light microscopy images of recoded strain A2.

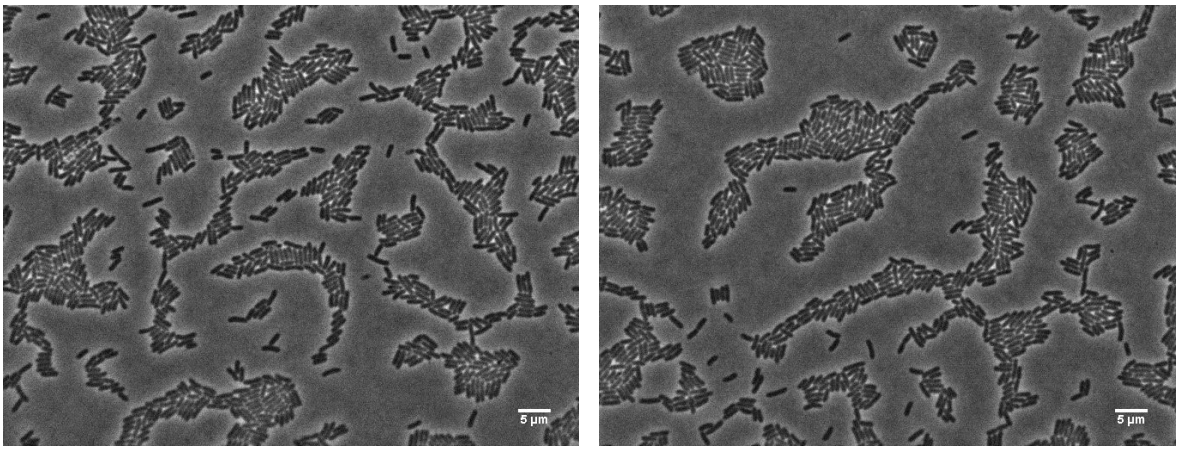

Figure S7.2.5. Light microscopy images of recoded strain A3.

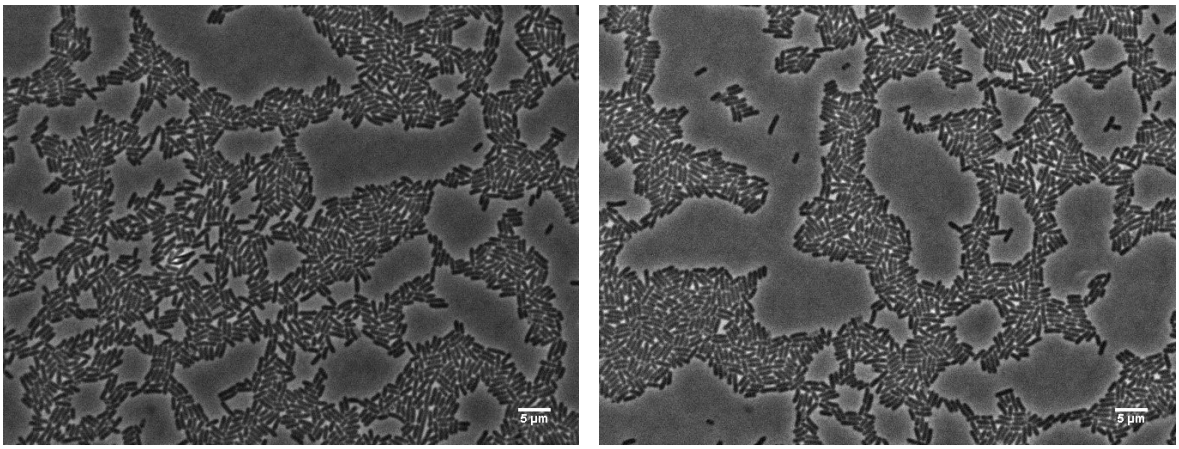

Figure S7.2.6. Light microscopy images of recoded strain A4.

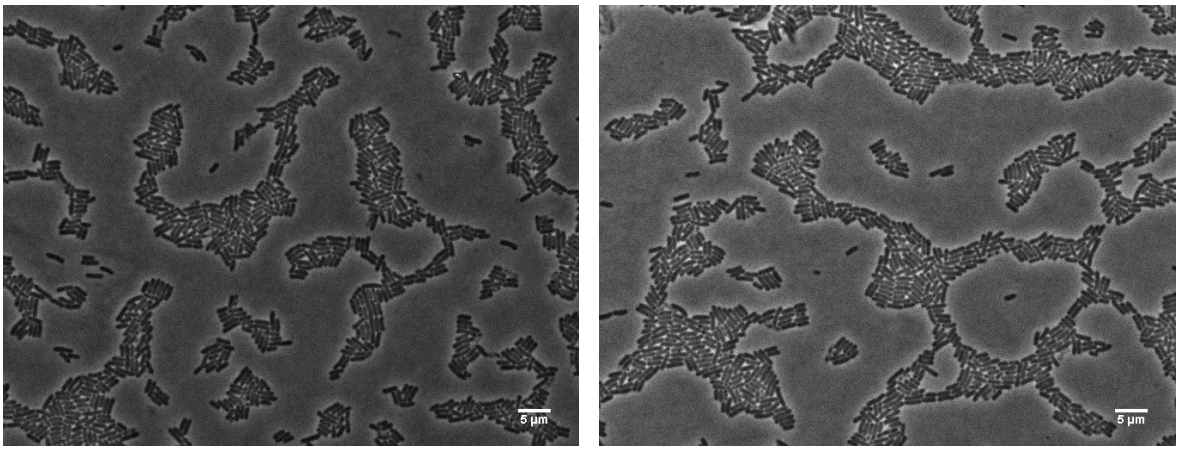

Figure S7.2.7. Light microscopy images of recoded strain A5.

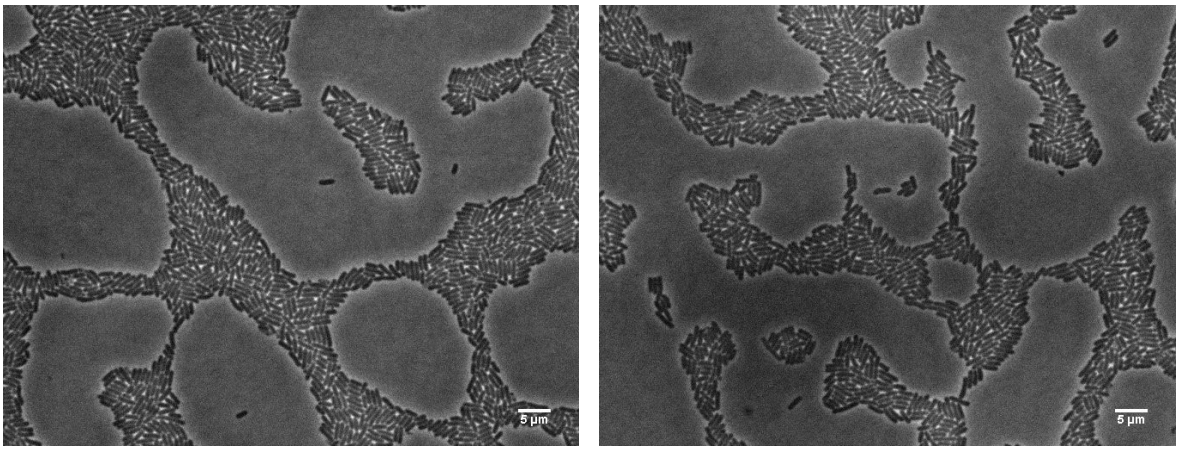

Figure S7.2.8. Light microscopy images of recoded strain A6.

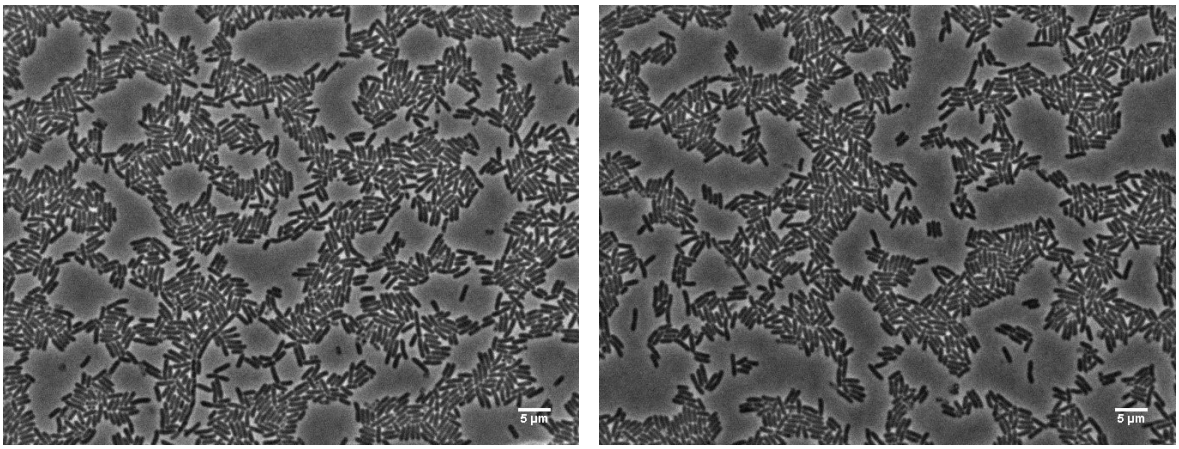

Figure S7.2.9. Light microscopy images of recoded strain A7.

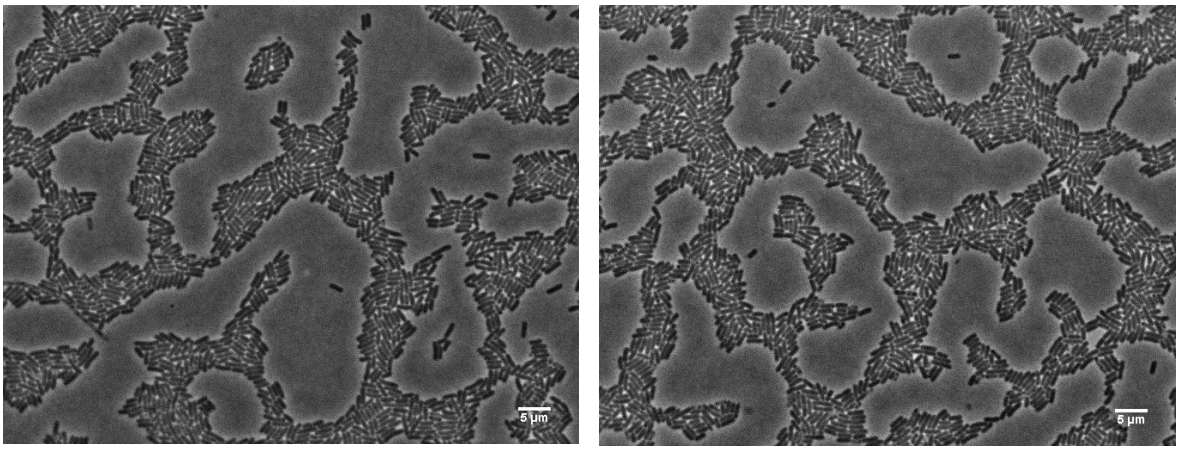

Figure S7.2.10. Light microscopy images of recoded strain A8.

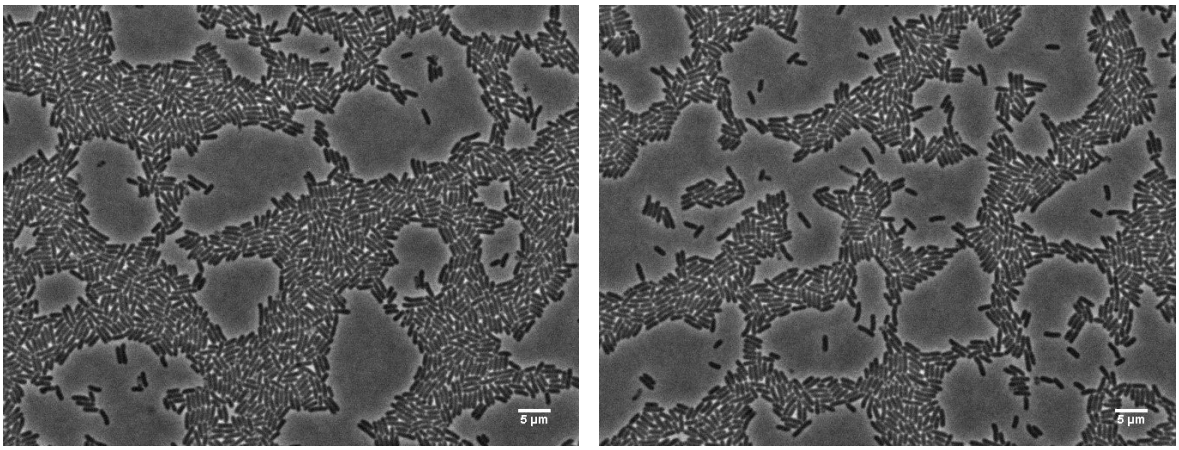

Figure S7.2.11. Light microscopy images of recoded strain A9.

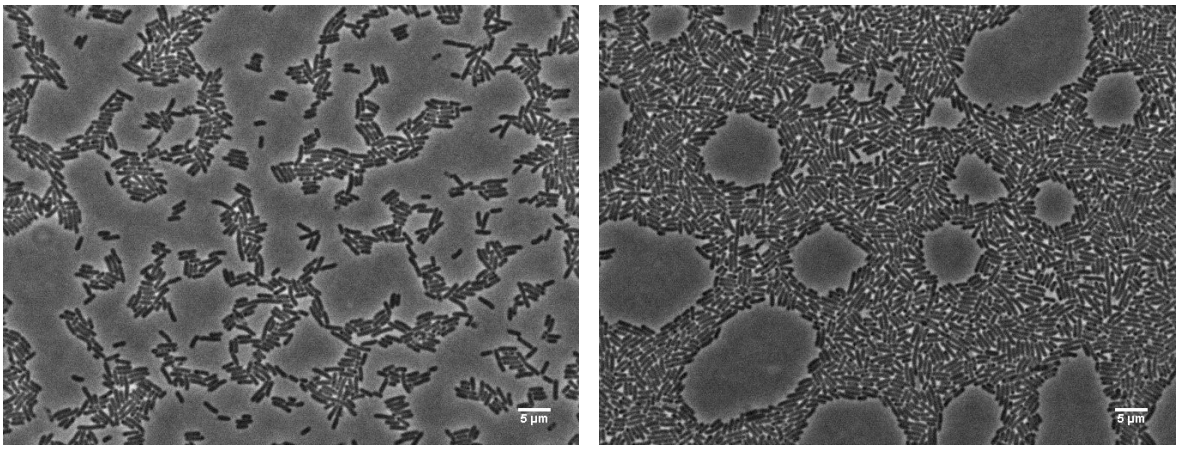

Figure S7.2.12. Light microscopy images of recoded strain A10.

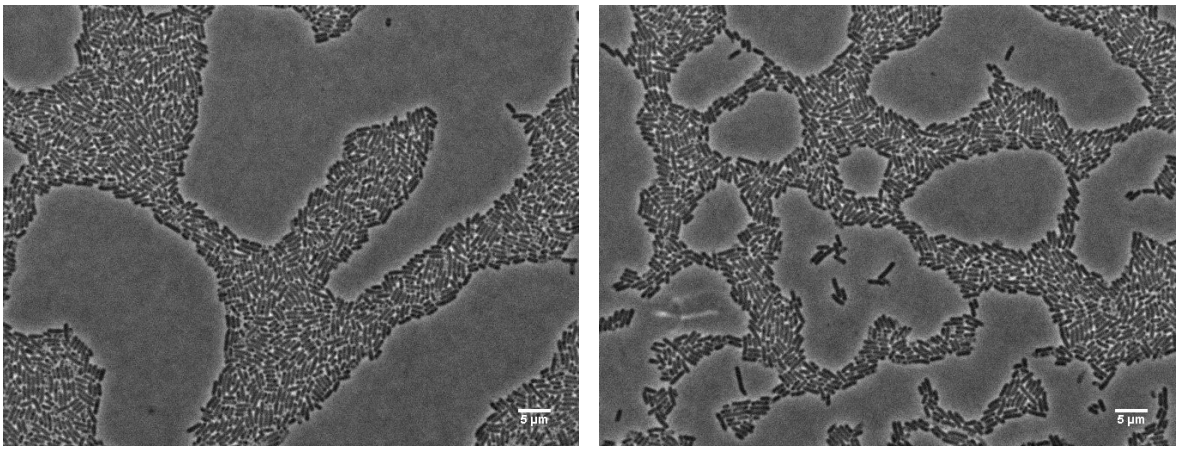

Figure S7.2.13. Light microscopy images of recoded strain A11.

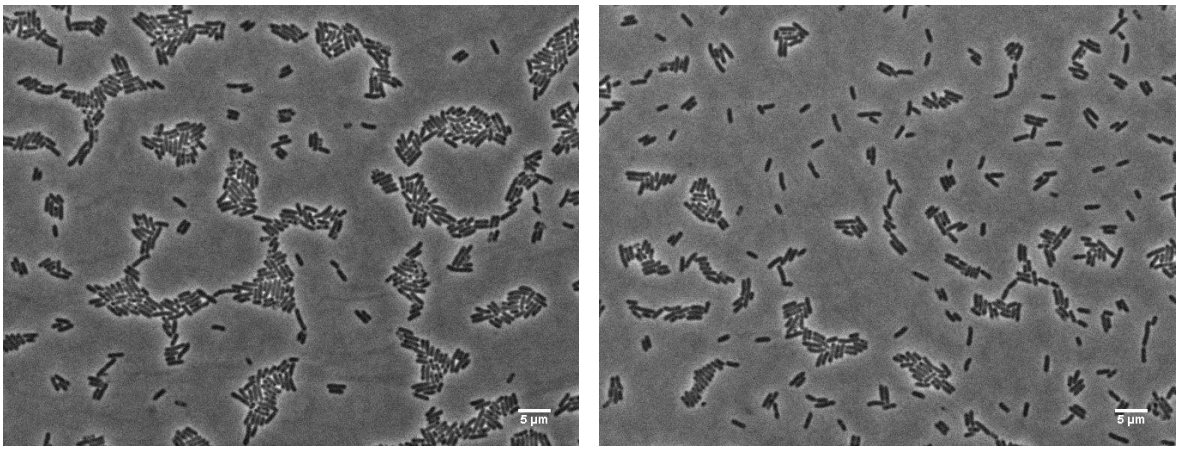

Figure S7.2.14. Light microscopy images of recoded strain A12.

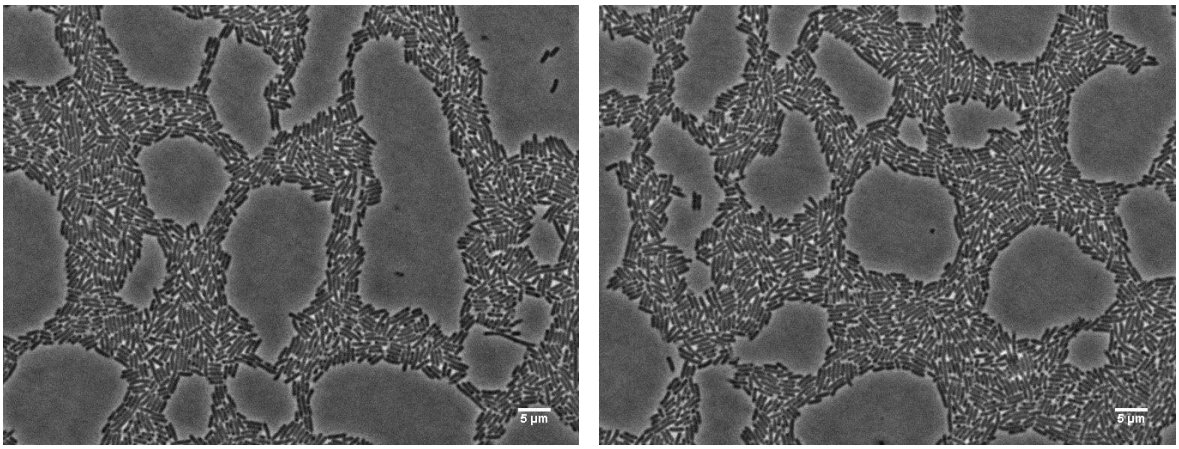

Figure S7.2.15. Light microscopy images of recoded strain A13.

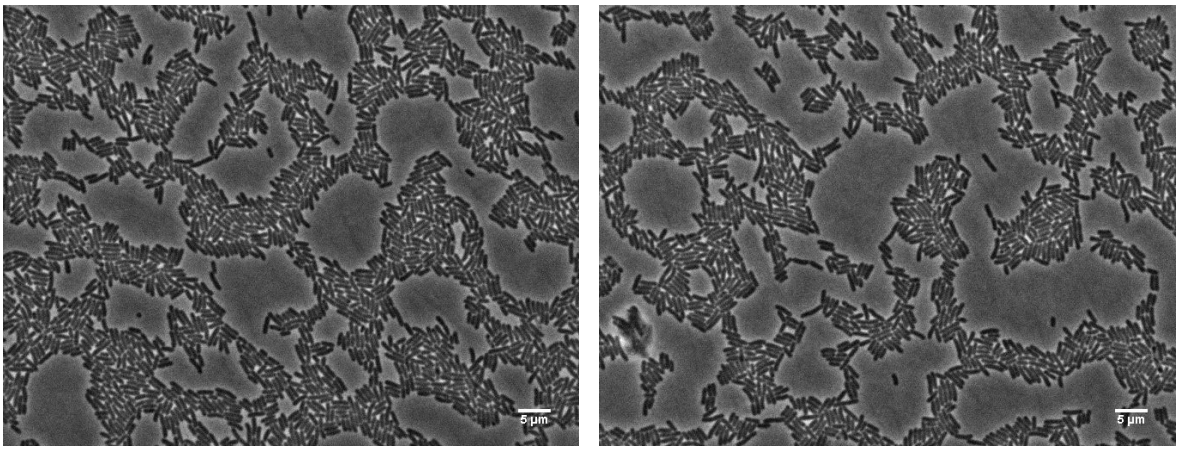

Figure S7.2.16. Light microscopy images of recoded strain B1.

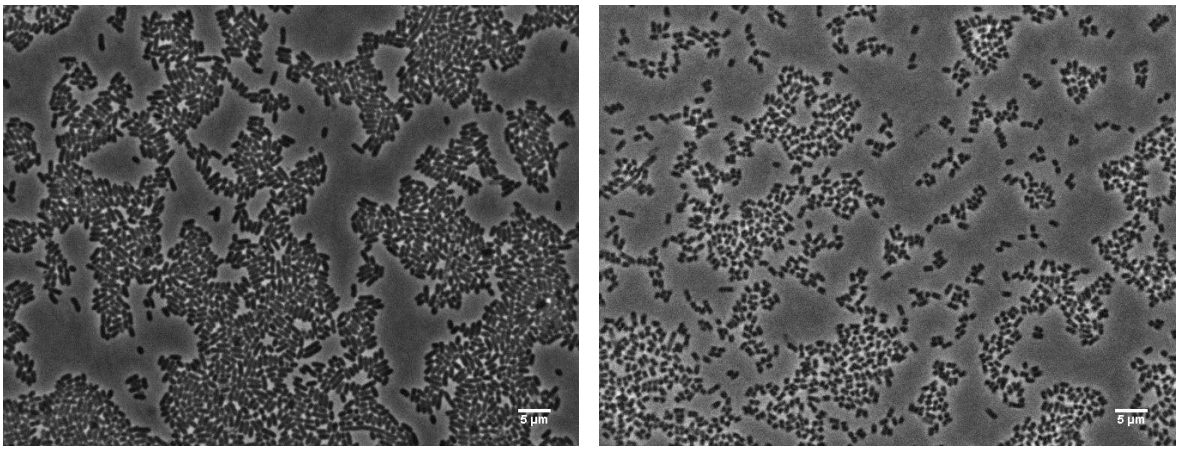

Figure S7.2.17. Light microscopy images of recoded strain B2.

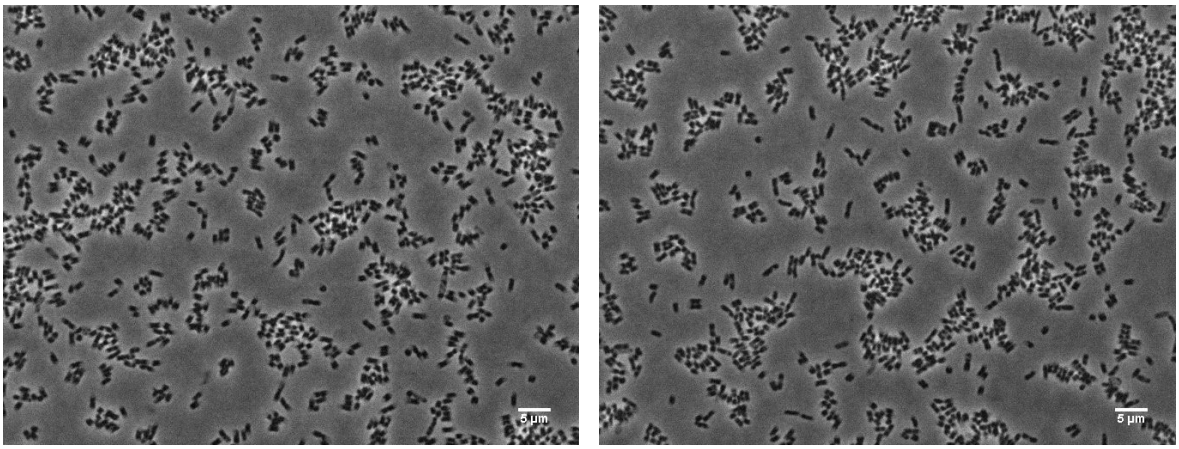

Figure S7.2.18. Light microscopy images of recoded strain B3.

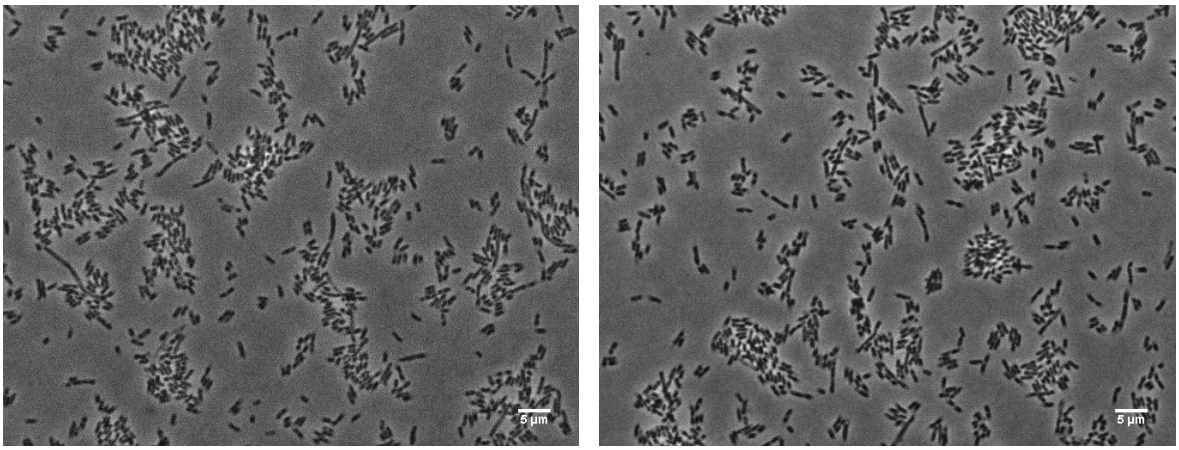

Figure S7.2.19. Light microscopy images of recoded strain A13-B3.

Chapter 3 supplementary material

Figure S1. Related to Figure 2: Sequences of essentializer degenerate library. Green = the antitoxin ORF and regulatory regions, Red = the toxin ORF and regulatory regions. Grey = the binding regions for cI and Cro, * indicating it has been modified to increase binding efficiency according to previously published work (Sarai & Takeda, 1989; Takeda et al., 1989).

Figure S2. Related to Figure 4: Fluorescence intensity for images in figure 4B. Images were partitioned into 5 section between streaked bacteria. A histogram of fluorescence intensity using arbitrary units is plotted for the strains containing different plasmids at a range of temperatures.

Figure S3. Related to Figure 5: Sequence of cryodeath degenerate library. Green = the antitoxin ORF and regulatory regions, Red = the toxin ORF and regulatory regions. Purple = toxin long

5' UTR. Blue = first 13 amino acids of *cspA* ORF, including the downstream box. Purple is the GGGGS linker that is present in all candidates except CD1 and CD11. Black annotations represent the regions that are deleted in CD01 and CD10.

Figure S4. Related to Figure 6: Survival ratio of candidates CD01 and CD10 in Dh10β at

30 ^{o}C, room temperature (~22 ^{o}C), and 15 ^{o}C. Data points are from average of 3 technical repeats.

Figure S5. Related to Figure 3 and Figure 6: A) Analysis of EE toxin only constructs after 140 generations of growth. Percentage of Colonies that retained the memory element after the assay outlined in figure 3A was performed. The toxin only constructs EE10 2 and EE11 3 were selected as having functional kill switches after growth in TB10 (Table S4). Six biological repeats of each were grown for 140 generations, along with 6 repeats of the toxin mutant control. **B)** Survival assay of CD10 toxin only constructs. To test if the antitoxin was required to maintain the stability of CD10, a construct that was identical to CD10 except with no antitoxin ORF was recombined into the DH10β using pKD46. Seven recombinant colonies were selected for analysis. Each was subjected to the survival assay (Figure 6A), with the non-permissible temperature of 22 ^{o}C. The seven constructs with sequences displayed in Table S8 were subjected to the survival assay outlined in figure 6A. All seven failed to display a cold sensitivity at 22 ^{o}C.

Figure S6. Related to Figure 1: A) Heat map of probability that for a given toxin-antitoxin pairing, the expression rate of the antitoxin is greater than or equal to the expression rate of the toxin. Assuming that transcription rate per generation is Poisson distributed with λ equal to the Antitoxin(λ) or Toxin(λ). Stochasticity due to transcription is expected to contribute much more to overall noise than stochasticity in translation, so the latter is ignored in this analysis. In addition, we assumed for simplicity that equal amounts of toxin and antitoxin transcription result in equal amounts of product, but in reality antitoxins are generally unstable and need to be expressed at a higher level to achieve equal amounts of product during steady-state expression. **B)** Graph displaying the values at which the expression rate of the antitoxin is greater than or equal to the expression rate of the toxin in any given generation, with given probability thresholds. These probability thresholds are directly translated into selection coefficients (s) which are displayed for each region of the graph. For example, if the probability that the [toxin] > [antitoxin] for any particular generation, then there will be a selection coefficient of 0.1 against the presence of the kill switch. It is important to note that for high transcription rates, the ratio between the maximal 'non-evolving' ratio and the minimal 'lethal' ratio can be as little as about threefold.

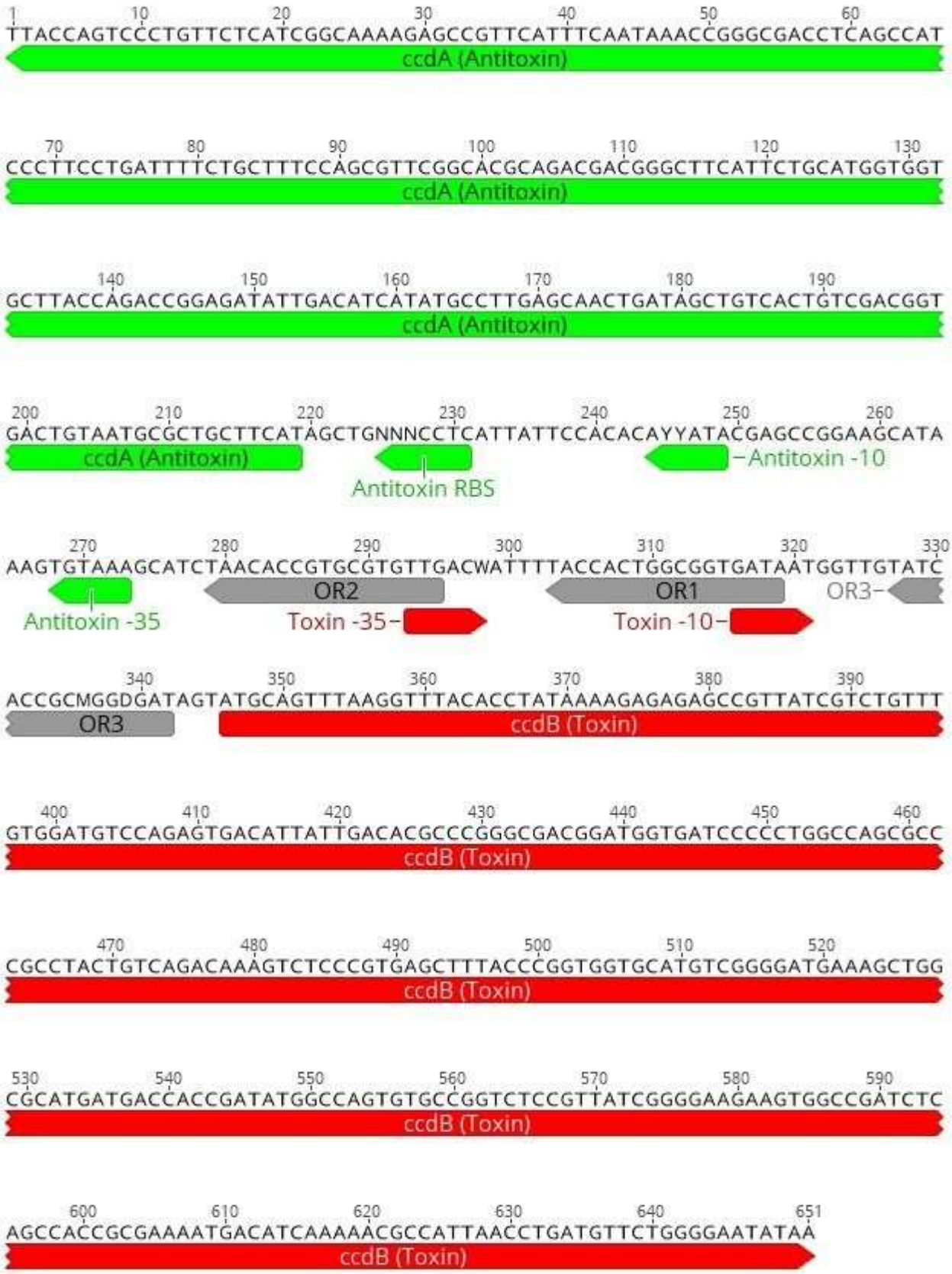
1 10 20 30 40 50 60
TTACCAGTCCCTGTTCTCATCGGCAAAAGAGCCGTTCATTTCAATAAACCGGGCGACCTCAGCCAT
ccdA (Antitoxin)
70 80 90 100 110 120 130
CCCTTCCTGATTTTCTGCTTTCCAGCGTTCGGCACGCAGACGACGGGCTTCATTCTGCATGGTGGT
ccdA (Antitoxin)
140 150 160 170 180 190
GCTTACCAGACCGGAGATATTGACATCATATGCCTTGAGCAACTGATAGCTGTCACTGTCGACGGT
ccdA (Antitoxin)
200 210 220 230 240 250 260
GACTGTAATGCGCTGCTTCATAGCTGNNNCCTCATTATTCCACACAYYATACGAGCCGGAAGCATA
ccdA (Antitoxin)
Antitoxin RBS
Antitoxin -10
270 280 290 300 310 320 330
AAGTGTAAAGCATCTAACACCGTGCGTGTTGACWATTTTACCACTGGCGGTGATAATGGTTGTATC
Antitoxin -35
OR2
Toxin -35
OR1
Toxin -10
OR3
340 350 360 370 380 390
ACCGCMGGDGATAGTATGCAGTTTAAGGTTTACACCTATAAAAGAGAGAGCCGTTATCGTCTGTTT
OR3
ccdB (Toxin)
400 410 420 430 440 450 460
GTGGATGTCCAGAGTGACATTATTGACACGCCCGGGCGACGGATGGTGATCCCCCTGGCCAGCGCC
ccdB (Toxin)
470 480 490 500 510 520
CGCCTACTGTCAGACAAAGTCTCCCGTGAGCTTTACCCGGTGGTGCATGTCGGGGATGAAAGCTGG
ccdB (Toxin)
530 540 550 560 570 580 590
CGCATGATGACCACCGATATGGCCAGTGTGCCGGTCTCCGTTATCGGGGAAGAAGTGGCCGATCTC
ccdB (Toxin)
600 610 620 630 640 651
AGCCACCGCGAAAATGACATCAAAAACGCCATTAACCTGATGTTCTGGGGAATATAA
ccdB (Toxin)

Figure S1

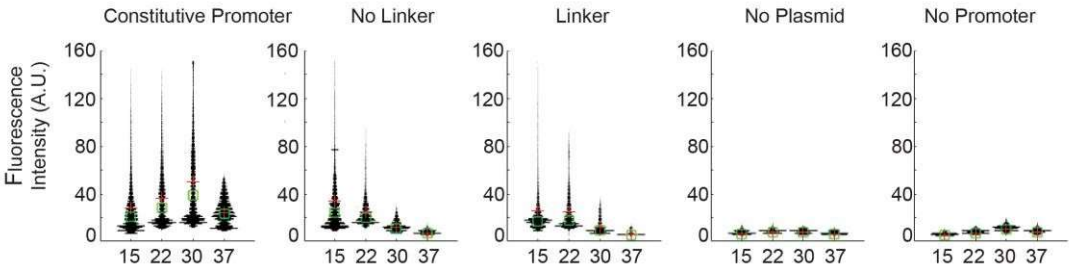

Figure S2

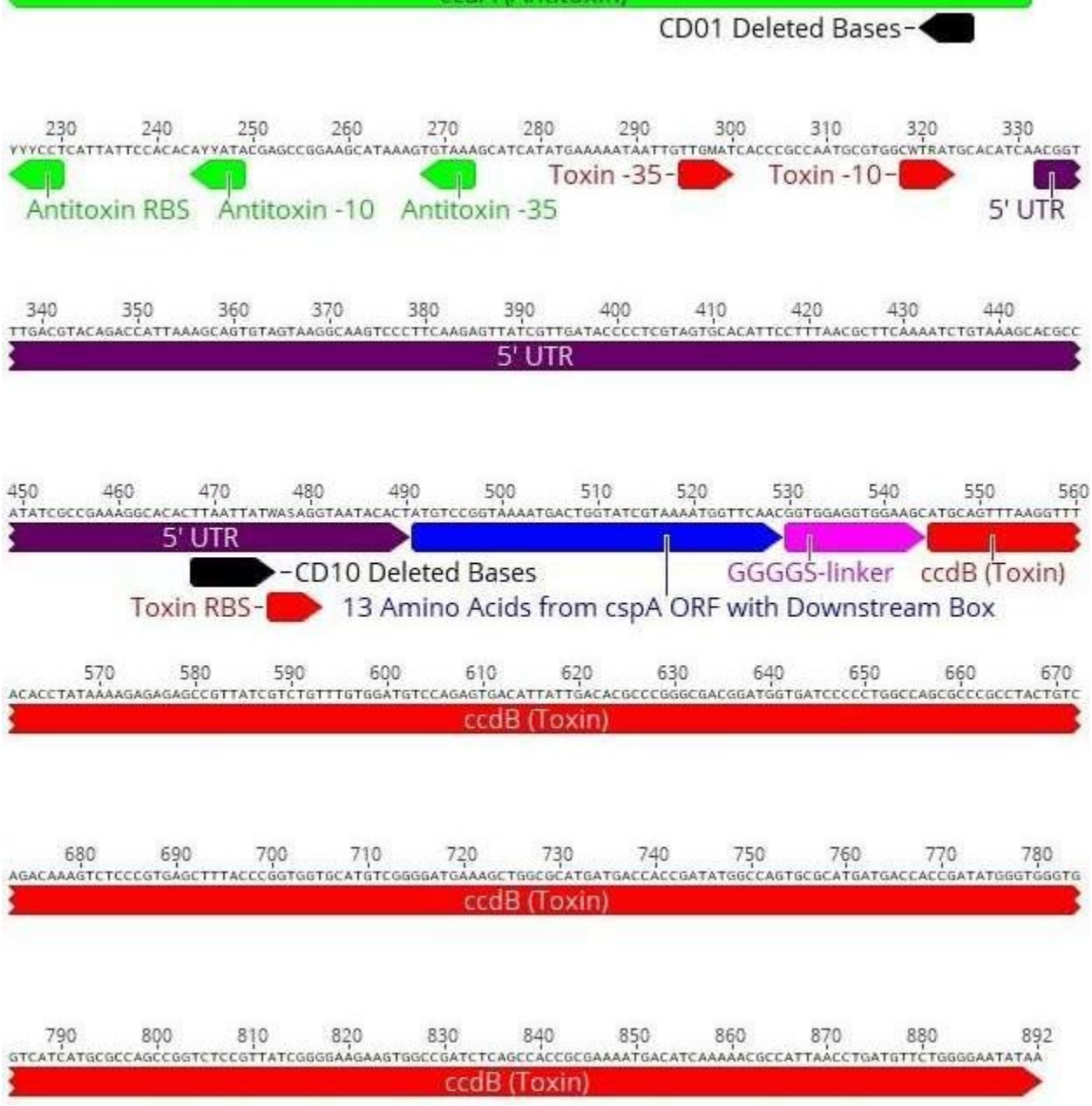
CD01 Deleted Bases
230 240 250 260 270 280 290 300 310 320 330
YYYCCTCATTATTCCACACAYYATACGAGCCGGAAGCATAAAGTGTAAAGCATCATATGAAAAATAATTGTTGMATCACCCGCCAATGCGTGGCWTRATGCACATCAACGGT
Antitoxin RBS
Antitoxin -10
Antitoxin -35
Toxin -35-
Toxin -10-
5' UTR
340 350 360 370 380 390 400 410 420 430 440
TTGACGTACAGACCATTAAAGCAGTGTAGTAAGGCAAGTCCCTTCAAGAGTTATCGTTGATACCCCTCGTAGTGCACATTCCTTTAACGCTTCAAAATCTGTAAAGCACGCC
5' UTR
450 460 470 480 490 500 510 520 530 540 550 560
ATATCGCCGAAAGGCACACTTAATTATWASAGGTAATACACTATGTCCGGTAAAATGACTGGTATCGTAAAATGGTTCAACGGTGGAGGTGGAAGCATGCAGTTTAAGGTTT
5' UTR
-CD10 Deleted Bases
GGGGS-linker
ccdB (Toxin)
Toxin RBS-
13 Amino Acids from cspA ORF with Downstream Box
570 580 590 600 610 620 630 640 650 660 670
ACACCTATAAAAGAGAGAGCCGTTATCGTCTGTTTGTGGATGTCCAGAGTGACATTATTGACACGCCCGGGCGACGGATGGTGATCCCCCTGGCCAGCGCCCGCCTACTGTC
ccdB (Toxin)
680 690 700 710 720 730 740 750 760 770 780
AGACAAAGTCTCCCGTGAGCTTTACCCGGTGGTGCATGTCGGGGATGAAAGCTGGCGCATGATGACCACCGATATGGCCAGTGCGCATGATGACCACCGATATGGGTGGGTG
ccdB (Toxin)
790 800 810 820 830 840 850 860 870 880 892
GTCATCATGCGCCAGCCGGTCTCCGTTATCGGGGAAGAAGTGGCCGATCTCAGCCACCGCGAAAATGACATCAAAAACGCCATTAACCTGATGTTCTGGGGAATATAA
ccdB (Toxin)

Figure S3

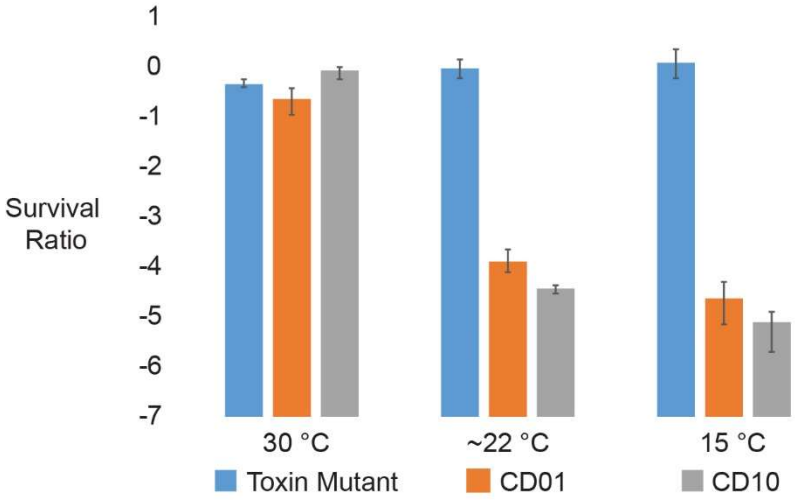
1
0
-1
-2
-3
-4
-5
-6
-7
Survival Ratio
30 °C
~22 °C
15 °C
Toxin Mutant
CD01
CD10

Figure S4

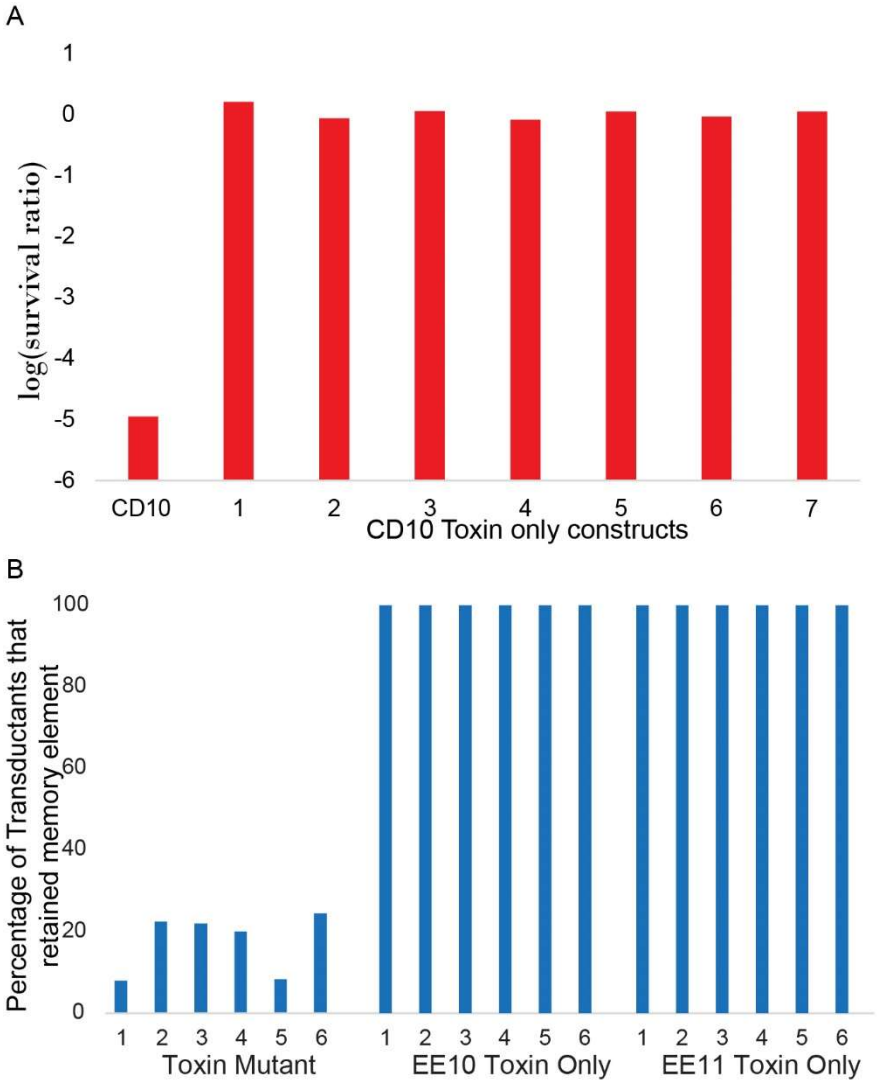
A
1
0
-1
-2
-3
-4
-5
-6
log(survival ratio)
CD10
1
2
3
4
5
6
7
CD10 Toxin only constructs
B
100
80
60
40
20
0
Percentage of Transductants that retained memory element
1 2 3 4 5 6
Toxin Mutant
1 2 3 4 5 6
EE10 Toxin Only
1 2 3 4 5 6
EE11 Toxin Only

Figure S5

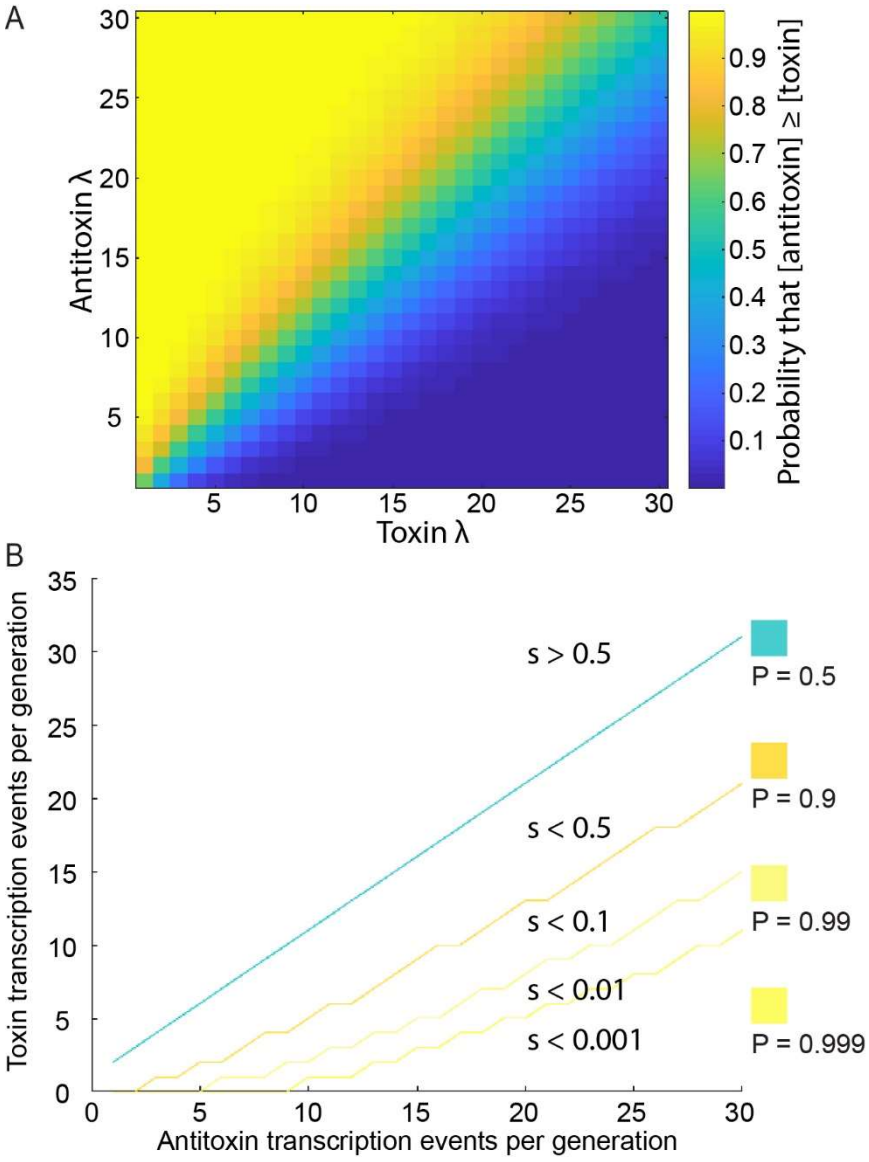
A
Antitoxin λ
Toxin λ
Probability that [antitoxin] ≥ [toxin]
B
Toxin transcription events per generation
Antitoxin transcription events per generation
s > 0.5
s < 0.5
s < 0.1
s < 0.01
s < 0.001
P = 0.5
P = 0.9
P = 0.99
P = 0.999

Figure S6

Table S1. Related to Figure 2: Raw data from tests 1-4 of flowchart in Figure 2E.

Table S2. Related to Figure 3: Results of growing toxin only versions of EE10 and EE11 in TB10 and PAS132 backgrounds. Failure to grow in TB10 at 42^{o}C implies a functional essentializer element. Capacity to select for the memory element was determined using the assay outlined in Figure 3A.

Table S3. Related to Figure 3 and Figure 6: Whole genome sequencing of the non-temperature sensitive CD08 strains and CD10 strains that grew at 22 ^{o}C. For CD08, mutations were identified in only one of two strains sequenced. Two mutations were identified in this strain. For CD10 strains, mutations were identified in 3 out of 5 strains sequenced.

Table S4. Related to Figure 3 and Figure 6: Sequences of regulatory region and ORF of ccdB CD10 toxin only constructs: Spaces indicate were a deletion has occurred compared to parental CD10. All seven CD10 toxin only constructs have frame shift mutations within the ORF of *ccdB*.

	Inhibited growth at	No. transductants from MG1655 lysate into	No. transducatants from candidates lysate into	No. transducatants from candidates lysate into PAS132
EE01	Yes	0	0	7
EE02	Yes	0	0	6
EE03	Yes	0	0	7
EE04	Yes	0	0	2
EE05	Yes	0	0	12
EE06	Yes	0	0	10
EE07	Yes	0	0	10
EE08	Yes	0	0	11
EE09	Yes	0	0	16
EE10	Yes	0	0	7
EE11	Yes	0	0	3
EE12	Yes	22	0	8
EE13	Yes	1	0	10
EE14	Yes	0	1	10
EE15	Yes	8	0	10
EE16	Yes	13	0	5
EE17	Yes	20	2	9
EE18	Yes	0	1	19
EE19	Yes	216	0	6
EE20	Yes	7	0	7
EE21	Yes	97	0	5
EE22	Yes	13	7	13
EE23	Yes	10	0	4
EE24	No	6	1	3
EE25	No	na	na	na
EE26	No	30	0	5
EE27	No	na	na	na
EE28	No	12	7	11
EE29	No	na	na	na
EE30	No	11	25	14
EE31	No	0	8	7
NL	No	142	5	13
TM	No	48	16	10

Table S3

	Failed to grow at 42°C in TB10	Fully Selected for the Presence of the Memory Element in PAS132
EE10 toxin only 1	Yes	No
EE10 toxin only 2	Yes	Yes
EE10 toxin only 3	No	No
EE10 toxin only 4	No	No
EE10 toxin only 5	Yes	No
EE10 toxin only 6	Yes	Yes
EE10 toxin only 7	Yes	Yes
EE10 toxin only 8	Yes	Yes
EE11 toxin only 1	Yes	Yes
EE11 toxin only 2	Yes	Yes
EE11 toxin only 3	Yes	Yes
EE11 toxin only 4	No	No
EE11 toxin only 5	Yes	Yes
EE11 toxin only 6	No	No
EE11 toxin only 7	Yes	Yes

Table S4

CD08 Non Temperature Sensitive Strains	Gene	Alteration
1		
2	5' UTR of CspA	G→T
2	*gtrS*	529 bp deletion
CD10 Escapees	Gene	Alteration
1		
2	5' UTR of CcdB	215 → 64 bp
3	*yedN*	+4 bp
4	promoter of *ccdB*	+9 bp
5		

Table S6

CD10 original	ATATGAAAAATAATTGTTGCATCACCCGCCAATGCGTGGCTTGATGCACATCAACGGTTTGACGTACAGACCATTAAAGCAGTGTAGTAAGGCAAGTCCCTTCAAGAGTTATCGTTGATACCCCTCGTAGT GCACATTCCTTTAACGCTTCAAAATCTGTAAAGCACCGCCATATCGCCGAAAGGCACACAGAGGTAATACACTATGTCCGGTAAAATGACTGGTATCGTAAAATGGTTCAACGGTGGAGGTGGAAGCATGCA GTTTAAGGTTTACACCTATAAAAGAGAGAGCCGTTATCGTCTGTTT-TGGATGTCCAGAGTGACATTATTGACACGCCCGGGCGACGGATGGTGATCCCCCTGGCCAGCGCCCGCCTACTGTCAGACAAAG TCTCCCGTGAGCTTTACCCGGTGGTGCATGTCGGGGATGAAAGCTGGCGCATGATGATGACCACCGATATGGCCAGTGTGCCGGTC-TCCGTTATGGGGGAAGAAGTGGCCGATCTCAGCCACCGCGGAAAATGA CATCAAAAACGCCATTAACCTGATGTTCTGGGGAATATAA
CD10 toxin only 7	ATATGAAAAATAATTGTTGCATCACCCGCCAATGCGTGGCTTGATGCACATCAACGGTTTGACGTACAGACCATTAAAGCAGTGTAGTAAGGCAAGTCCCTTCAAGAGTTATCGTTGATACCCCTCGTAGT GCACATTCCTTTAACGCTTCAAAATCTGTAAAGCACCGCCATATCGCCGAAAGGCACACAGAGGTAATACACTATGTCCGGTAAAATGACTGGTATCGTAAAATGGTTCAACGGTGGAGGTGGAAGCATGCA GTTTAAGGTTTACACCTATAAAAGAGAGAGCCGTTATCGTCTGTTTGTGGATGTCCAGAGTGACATTATTTGACACGCCCGGGCGACGGATGGTGATCCCCCTGGCCAGCGCCCGCCTACTGTCAGACAAAG TCTCCCGTGAGCTTTACCCGGTGGTGCATGTCGGGG-ATGAAAGCTGGCGCATGATGATGACCACCGATATGGCCAGTGTGCCGGTC-TCCGTTATGGGGGAAGAAGTGGCCGATCTCAGCCACCGCGGAAAATGA CATCAAAAACGCCATTAACCTGATGTTCTGGGGAATATAA
CD10 toxin only 6	ATATGAAAAATAATTGTTGCATCACCCGCCAATGCGTGGCTTGATGCACATCAACGGTTTGACGTACAGACCATTAAAGCAGTGTAGTAAGGCAAGTCCCTTCAAGAGTTATCGTTGATACCCCTCGTAGT GCACATTCCTTTAACGCTTCAAAATCTGTAAAGCACCGCCATATCGCCGAAAGGCACACAGAGGTAATACACTATGTCCGGTAAAATGACTGGTATCGTAAAATGGTTCAACGGTGGAGGTGGAAGCATGCA GTTTAAGGTTTACACCTATAAAAGAGAGAGCCGTTATCGTCTGTTTGTGGATGTCCAGAGTGACATTATTTGACACGCCCGGGCGACGGATGGTGATCCCCCTGGCCAGCGCCCGCCTACTGTCAGACAAAG TCTCCCGTGAGCTTTACCCGGTGGTGCATGTCGGGGATGAAAGCTGGCGCATGATGATGACCACCGATATGGCCAGTGTGCCGGTC-TCCGTTATGGGGGAAGAAGTGGCCGATCTCAGCCACCGCGGAAAATGA CATCAAAAACGCCATTAACCTGATGTTCTGGGGA-TATAA
CD10 toxin only 5	ATATGAAAAATAATTGTTGCATCACCCGCCAATGCGTGGCTTGATGCACATCAACGGTTTGACGTACAGACCATTAAAGCAGTGTAGTAAGGCAAGTCCCTTCAAGAGTTATCGTTGATACCCCTCGTAGT GCACATTCCTTTAACGCTTCAAAATCTGTAAAGCACCGCCATATCGCCGAAAGGCACACAGAGGTAATACACTATGTCCGGTAAAATGACTGGTATCGTAAAATGGTTCAACGGTGGAGGTGGAAGCATGCA GTTTAAGGTTTACACCTATAAAAGAGAGAGCCGTTATCGTCTGTTTGTGGATGTCCAGAGTGACATTATTTGACACGCCCGGGCGACGGATGGTGATCCCCCTGGCCAGCGCCCGCCTACTGTCAGACAAAG TCTCCC-GTGAGCTTTACCCGGTGGTGCATGTCGGGGATGAAAGCTGGCGCATGATGATGACCACCGATATGGCCAGTGTGCCGGTC-TCCGTTATGGGGGAAGAAGTGGCCGATCTCAGCCACCGCGGAAAATGA CATCAAAAACGCCATTAACCTGATGTTCTGGGGAATATAA
CD10 toxin only 4	ATATGAAAAATAATTGTTGCATCACCCGCCAATGCGTGGCTTGATGCACATCAACGGTTTGACGTACAGACCATTAAAGCAGTGTAGTAAGGCAAGTCCCTTCAAGAGTTATCGTTGATACCCCTCGTAGT GCACATTCCTTTAACGCTTCAAAATCTGTAAAGCACCGCCATATCGCCGAAAGGCACACAGAGGTAATACACTATGTCCGGTAAAATGACTGGTATCGTAAAATGGTTCAACGGTGGAGGTGGAAGCATGCA GTTTAAGGTTTACACCTATAAAAGAGAGAGCCGTTATCGTCTGTTTGTGGATGTCCAGAGT-ACATTATTGACACGCCCGGGCGACGGATGGTGATCCCCCTGGCCAGCGCCCGCCTACTGTCAGACAAAG TCTCCCGTGAGCTTTACCCGGTGGTGCATGTCGGGGATGAAAGCTGGCGCATGATGATGACCACCGATATGGCCAGTGTGCCGGTCTCCTCCGTTATCGGGGAAGAAGTGGCCGATCTCAGCCACCGCGGAAAATGA CATCAAAAACGCCATTAACCTGATGTTCTGGGGAATATAA
CD10 toxin only 3	ATATGAAAAATAATTGTTGCATCACCCGCCAATGCGTGGCTTGATGCACATCAACGGTTTGACGTACAGACCATTAAAGCAGTGTAGTAAGGCAAGTCCCTTCAAGAGTTATCGTTGATACCCCTCGTAGT GCACATTCCTTTAACGCTTCAAAATCTGTAAAGCACCGCCATATCGCCGAAAGGCACACAGAGGTAATACACTATGTCCGGTAAAATGACTGGTATCGTAAAATGGTTCAACGGTGGAGGTGGAAGCATGCA GTTTAAGGTTTACACCTATAAAAGAGAGAGCCGTTATCGTCTGTTTGTGGATGTCCAGAGTGACATTATTGACACGCCCGG-CGACGGATGGTGATCCCCCTGGCCAGCGCCCGCCTACTGTCAGACAAAG TCTCCCGTGAGCTTTACCCGGTGGTGCATGTCGGGGATGAAAGCTGGCGCATGATGATGACCACCGATATGGCCAGTGTGCCGGTC-TCCGTTATGGGGGAAGAAGTGGCCGATCTCAGCCACCGCGGAAAATGA CATCAAAAACGCCATTAACCTGATGTTCTGGGGAATATAA
CD10 toxin only 2	ATATGAAAAATAATTGTTGCATCACCCGCCAATGCGTGGCTTGATGCACATCAACGGTTTGACGTACAGACCATTAAAGCAGTGTAGTAAGGCAAGTCCCTTCAAGAGTTATCGTTGATACCCCTCGTAGT GCACATTCCTTTAACGCTTCAAAATCTGTAAAGCACCGCCATATCGCCGAAAGGCACACAGAGGTAATACACTATGTCCGGTAAAATGGTTCAACGGTGGAGGTGGAAGCATGCA GTTTAAGGTTTACACCTATAAAAGAGAGAGCCGTTATCGTCTGTTTGTGGATGTCCAGAGTGACATTATTTGACACGCCCGGGCGACGGATGGTGATCCCCCTGGCCAGCGCCCGCCTACTGTCAGACAAAG TCTCCCGTGAGCTTTACCCGGTGGTGCATGTCGGGGATGAAAGCTGGCGCATGATGATGACCACCGATATGGCCAGTGTGCCGGTC-TCCGTTATGGGGGAAGAAGTGGCCGATCTCAGCCACCGCGGAAAATGA CATCAAAAACGCCATTAACCTGATGTTCTGGGGAATATAA
CD10 toxin only 1	ATATGAAAAATAATTGTTGCATCACCCGCCAATGCGTGGCTTGATGCACATCAACGGTTTGACGTACAGACCATTAAAGCAGTGTAGTAAGGCAAGTCCCTTCAAGAGTTATCGTTGATACCCCTCGTAGT GCACATTCCTTTAACGCTTCAAAATCTGTAAAGCACCGCCATATCGCCGAAAGGCACACAGAGGTAATACACTATGTCCGGTAAAATGACTGGTATCGTAAAATGGTTCAACGGTGGAGGTGGAAGCATGCA GTTTAAGGTTTACACCTATAAAAGAGAGAGCCGTTATCGTCTGTTTGTGGATGTCCAGAGTGACATTATTTGACACGCCCGGGCGACGGATGGTGATCCCCCTGGCCAGCGCCCGCCTACTGTCAGACAAAG TCTCCC-GTGAGCTTTACCCGGTGGTGCATGTCGGGGATGAAAGCTGGCGCATGATGATGACCACCGATATGGCCAGTGTGCCGGTC-TCCGTTATGGGGGAAGAAGTGGCCGATCTCAGCCACCGCGGAAAATGA CATCAAAAACGCCATTAACCTGATGTTCTGGGGAATATAA

Table S8

Chapter 4 supplementary material

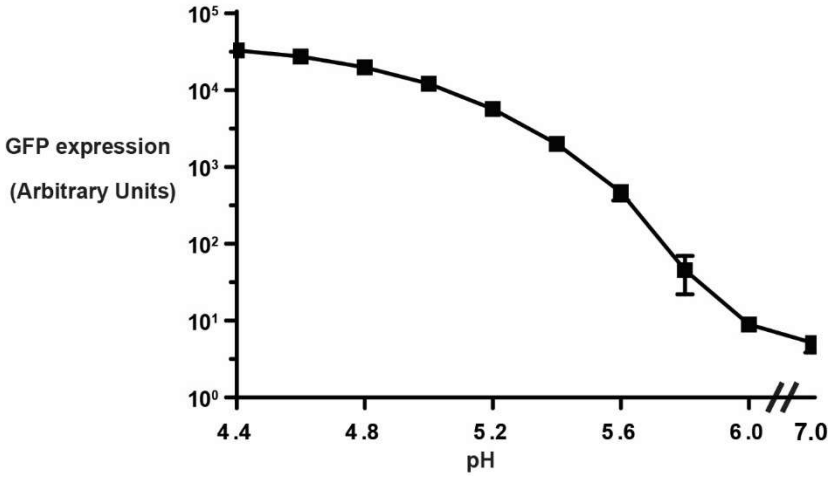

Figure S1. Related to Figure 1. Induction curve of P_{asr}AT09:GFP in M9 media with 1 g/L yeast extract buffered to a range of pH's. GFP fluorescence was measured using a flow cytometer after five hours of growth at 37 ^{o}C. data points are from three technical repeats, error bars are standard deviation.

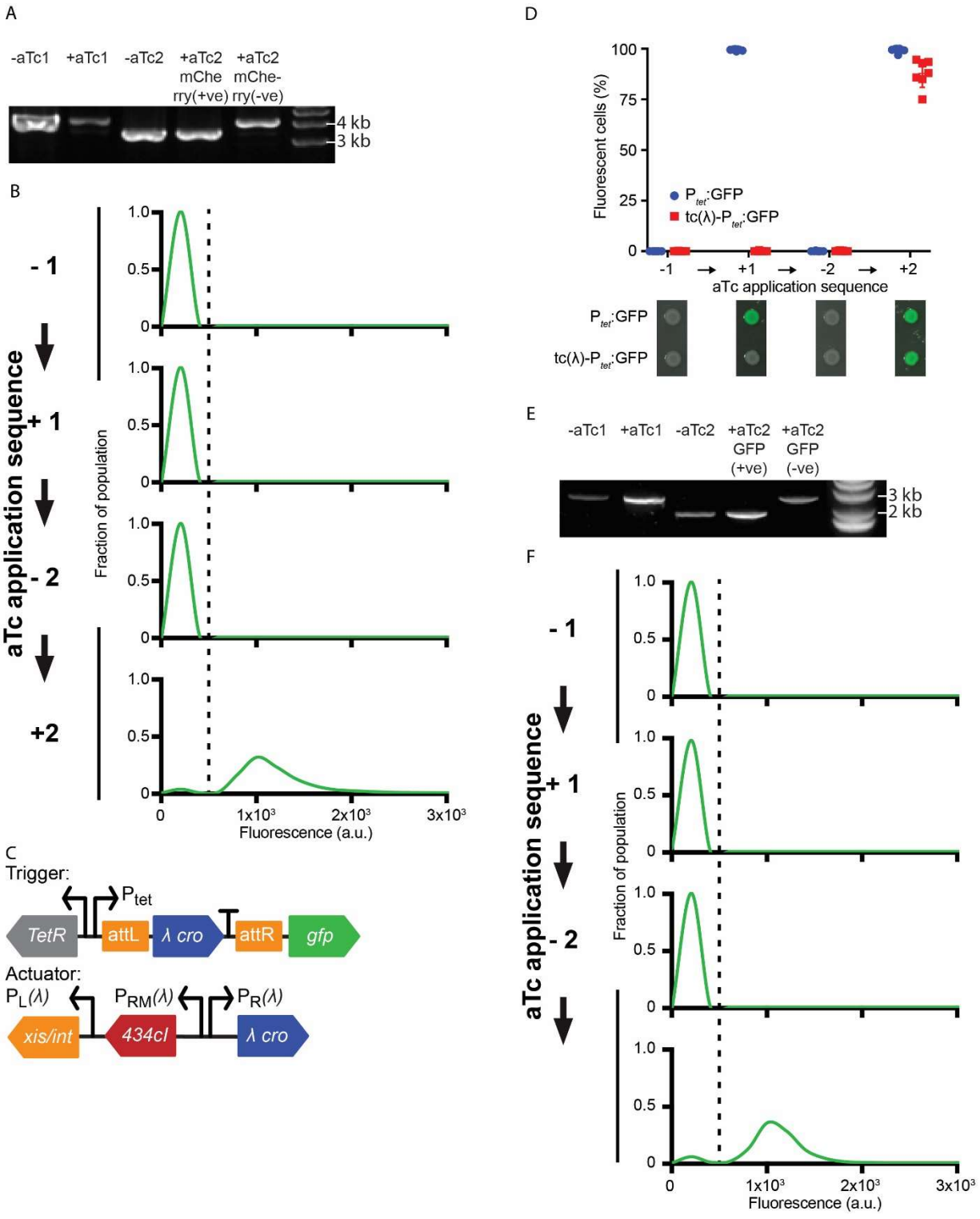
A
-aTc1
+aTc1
-aTc2
+aTc2 mChe rry(+ve)
+aTc2 mChe- rry(-ve)
4 kb
3 kb
B
aTc application sequence
- 1
+ 1
- 2
+2
Fraction of population
1.0
0.5
0
1x10³
2x10³
3x10³
Fluorescence (a.u.)
C
Trigger:
P tet
TetR
attL
λ cro
attR
gfp
Actuator:
P L(λ)
P RM(λ)
P R(λ)
xis/int
434cI
λ cro
D
Fluorescent cells (%)
100
75
50
25
0
P tet:GFP
tc(λ)-P tet:GFP
-1
+1
-2
+2
aTc application sequence
E
-aTc1
+aTc1
-aTc2
+aTc2 GFP (+ve)
+aTc2 GFP (-ve)
3 kb
2 kb
F
aTc application sequence
- 1
+ 1
- 2
Fraction of population
1.0
0.5
0
1x10³
2x10³
3x10³
Fluorescence (a.u.)

Figure S3. Related to Figure 3.

S3A) Agarose gel depicting bands from PCR amplification of the 434 two-counter trigger during a time course of aTc induction. At each step of the time course, genomic DNA from a colony of tc-P_{tet}:mCherry was used as a template for PCR at the Tn7 locus to check for excision of the trigger. (Full trigger length: 4.0 kb; excised trigger length: 3.3 kb.) Rightmost band corresponds to a single mCherry-negative colony picked from the +aTc2 plate.

S3B) Distribution of mCherry fluorescence intensity in a single culture of the tc-P_{tet}:mCherry, in which 97% of the population has successfully undergone trigger exicision as a result of the first aTc stimulus. The cutoff between mCherry negative and mCherry positive cells is indicated by the dotted line at ~100 a.u.

S3C) λ two-counter circuit with a GFP reporter (tc(λ)-P_{tetR}:GFP).

S3D) Response of tc(λ)-P_{tet}:GFP to an aTc induction time course with two stimulus pulses. aTc was applied and washed out for growth periods of at least 4 hours (+1 and +2 growth steps). A P_{tet}:GFP strain is shown as a control. Error bars represent SD of seven biological replicates. Spots of cultures on agar media (with or without aTc) are shown below at each step of the time course. Spot images consist of a GFP color overlay on a grayscale brightfield image.

S3E) Agarose gel depicting bands from PCR amplification of tc(λ)-P_{tet}:GFP at the araB/araC locus. At each step of the time course, genomic DNA from a colony of tc(λ)-P_{tet}:GFP was used as a template for PCR to check for excision of the trigger. (Full trigger length: 2.5 kb; excised trigger length: 1.8 kb.) Rightmost band corresponds to a single GFP-negative colony picked from the +aTc2 plate.

S3F) Distribution of GFP fluorescence intensity in a single culture of the tc(λ)-P_{tet}:GFP, in which 93% of the population has successfully undergone trigger excision as a result of the first aTc stimulus. The cutoff between GFP- and GFP+ cells is indicated by the dotted line at ~100 a.u.

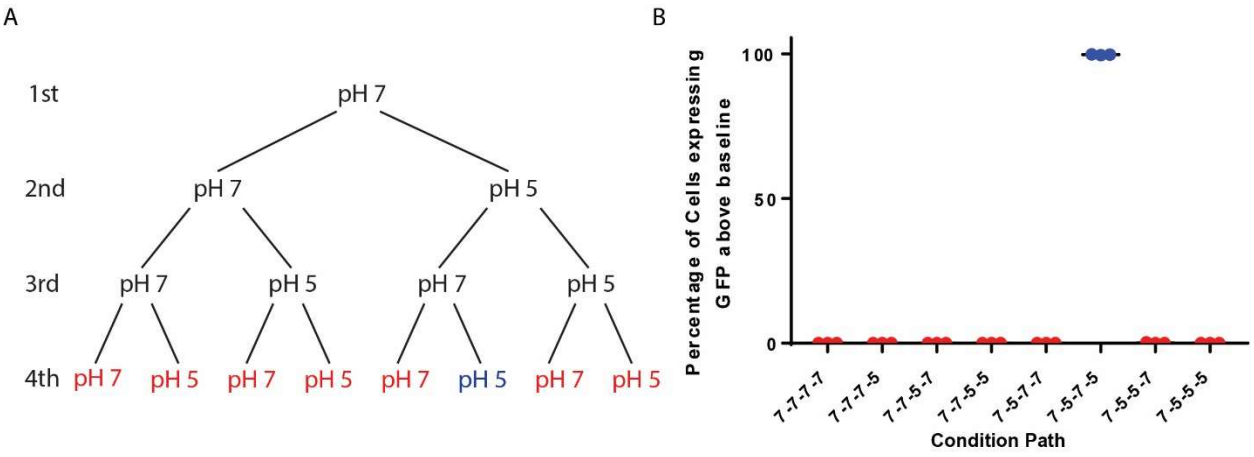

Figure S4. **A)** Tree of conditions for acid sensitive two count assay. each level of the tree represents a growth step, with each step after the first having one growth in pH 7 conditions and one in pH 5. On the final level, only the condition highlighted in blue would expect to express the reporter gene. All other conditions (red) would expect no expression. **B)** percentage of cells expressing GFP above that of a negative controls (wild-type and uninduced P_{asr}:GFP). Only cells exposed to two non-consecutive growth steps in pH 5 conditions showed a significant proportion of the population expressing GFP.

P_{asr}	tcaacgctgtaatttattcagcgtttgtacatatcgttacacgctgaaaccaaccactcacggaagtc tgccattcccagggatatagttatttcaacggccccgcagtggggttaaatgaaaaaacaaattgagg gtatgaca

Table S1. related to Figure 1.

Sequence of promoter and ribosome binding site of P_{asr} ordered from SGI DNA.

acidTRP degenerate library	5’-cgcttaatgctttccagatggcaacatgtgctggcagagcttcacgcagcgtatcgatatac Agctgttcttcattaccatgcagttgcactgccgccagttagaaaaactcatcgagcatcaagtg Aaactgcaatttattcatatcaggattatcaataccatattttgaaaaagccgtttctgtaatg aaggagaaaactcaccgaggcagttccataggatggcaagatcctggtatcggtctgcgattccg actcgtccaacatcaatacaacctattaatttcccctcgtcaaaaataaggttatcaagtgagaa atcaccatgagtgacgactgaatccggtgagaatggcaaaagcttatgcatttctttccagactt gttcaacaggccagccattacgctcgtcatcaaaatcactcgcaccaaccaaaccgttattcatt cgtgattgcgcctgagcgagacgaaatacgcgatcgccgttaaaaggacaattacaaacaggaat cgaatgcaaccggcgcaggaacactgccagcgcatcaacaatattttcacctgaatcaggatatt cttctaatacctggaatgctgttttccctgggatcgcagtggtgagtaaccatgcatcatcagga gtacggataaaatgcttgatggtcggaagaggcataaattccgtcagccagtttagcctgaccat ctcatctgtaacatcattggcaacgctacctttgccatgtttcagaaacaactctggcgcatcgg gcttcccatacaatcgatagattgtcgcacctgattgcccgacattatcgcgagcccatttatac ccatataaatcagcatccatgttggaatttaatcgcggcctcgagcaagacgtttcccgttgaat atggctcatatgtttttcctccttatgttaagcttactcagtcatcatatcgttgtacgagtaat aagtcaaaaaaaaactcccacatgggagttttatcggttaaccagttccttgttggtggagtcca gggtgtcaaacagggatgcaaattcagcatccagcgccgctttttgtaggcttcgaaagtagcc ttgctgacaattactgctggctcacggcctctgcgggtgatttcaacctcttccccggcttcaac attgttgagcacttcagaaaggttgccgcgcgcggtacggaagttaatggattgcatagctgYYY cctcattattccacacaYYatacgagccggaagcataaagtgtaaagtcaacgctgtaatttatt cagcgtttgtacatatcgttacacgctgaaaccaaccactcacggaagtctgccattcccagYga taKaRttatttcaacggccccgcagtggggttaaatgaaaaaacaaattgWgRgtatgacaatga ggcatatatcaccggaagaacttattgcgcttcatgatgcgaatataagccgctacggcggcctg ccgggaatgtctgatccgggtagggcagaggccattatcgggagagttcaggccagagttgccta cgaagagatcaccgacctttttcgaagtctccgccacctacctggtggctacagcgagagggcata tattcaatgatgccaataagcgtaccgcgctaaacagtgcgctgctatttctacgccgtaacggg gtgcaggtatttgattcacctgaactggcagaccttactgtaggagctgcgactggcgagatatc tgtatcttctgtcgccgacacgttacgtagattgtatggttctgcggagtagacgcaacgcggtg atgtcgcaacaaaaatcaacccaccgtaaatcgcgcccgcgtcataagctgctttagcatcttgc ccacgcgtcaggccacatactt-3’

Table S2, related to Figure 1.

Full sequence of degenerate acidTRP library ordered from SGI DNA. Degenerate bases are capitalised.

	Y	Y	Y	Y	Y	Y	K	R	W	R	Notes
acidTRP-TM	T	T	T	T	C	C	G	G	A	A	
acidTRP-01	T	C	C	C	C	C	G	G	A	G	
acidTRP-02	T	T	T	C	C	T	T	G	A	A	
acidTRP-03	C	T	T	C	T	C	G	G	A	A	
acidTRP-04	T	C	C	C	C	T	G	G	A	A	
acidTRP-05	C	T	C	T	T	T	G	A	A	G	A 1278 deleted
acidTRP-06	C	T	C	T	C	T	T	G	A	G	
acidTRP-07	T	C	C	C	C	T	T	A	A	G	T 1346 deleted
acidTRP-08	T	T	C	C	C	C	T	G	T	A	
acidTRP-09	T	T	C	T	C	T	G	G	A	G	
acidTRP-10	T	T	C	C	C	C	T	A	A	A	
acidTRP-11	T	T	C	C	T	C	T	G	T	G	
acidTRP-12	T	T	T	C	C	T	G	A	A	A	
acidTRP-13	T	T	T	C	C	T	T	A	A	A	

Table S3. Related to Figure 1. Identity of degenerate bases for acidTRP candidates reported on in Figure 1C. addition modifications are described under notes, with the bp position referencing the sequence in Table S2.

tc-acidTRP-09	5′– aatgcggtgagcatcacatcaccacaattcagcaaattgtgaacatcatcacgttcatctttccct ggttgccaatggcccattttcctgtcagtaacgaaaaaaaactcccacatgggagttttatcggtt aaccagttccttgttggtggagtccagggtgtcaaacagggatgcaaattcagcatccagcgccgc ttttttgtaggcttcgaaagtagccttgctgacaattactgctggctcacggcctctgcgggtgat ttcaacctcttccccggcttcaacattgttgagcacttcagaaaggttgccgcgcgcggtacggaa gttaatggattgcatagctgttccctcattattccacacatcatacgagccggaagcataaagtgt aaagggtcaacgctgtaatttattcagcgtttgtacatatcgttacacgctgaaaccaaccactca cggaagtctgccattcccagtgatagagttatttcaacggccccgcagtggggttaaatgaaaacc tgctttttatactaagttggcattataaaaaagcattgcttatcaatttgttgcaacgaacaggt cactatcagtcaaaataaaatcattatttgatttttgttgatggaggcgatatgcaaactctttct gaacgcctcaagaagaggcgaattgcgttaaaaatgacgcaaaccgaactggcaaccaaagccggt gttaaacagcaatcaattcaactgattgaagctggagtaaccaagcgaccgcgcttcttgtttgag attgctatggcgcttaactgtgatccggtttggttacagtacggaactaaacgcggtaaagccgct taaccaggcatcaaataaaacgaaaggctcagtcgaaagactgggcctttcgttttatctgttgtt tgtcggtgaacgctctctactagagtcacactggctcaccttcgggtgggcctttctgcgtttata aaattagcgcaagaagacaaaaatcaccttgcgctaatgctctgttacaggtcactaataccatct aagtagttgattcatagtgactgcatatgttgtgttttacagtattatgtagtctgtttttttatgc aaaatctaatttaatatattgatatttatatcattttacgtttctcgttcagctttttttatactaa cttgaaacaaattgagggtatgacaatgaggcatatatcaccggaagaacttattgcgcttcatga tgcgaatataagccgctacggcggcctgccgggaatgtctgatccgggtagggcagaggccattat cgggagagttcaggccagagttgcctacgaagagatcaccgacctttcgaagtctccgccaccta cctggtggctacagcgagagggcatatattcaatgatgccaataagcgtaccgcgctaaacagtgc gctgctatttctacgccgtaacggggtgcaggtatttgattcacctgaactggcagaccttactgt aggagctgcgactggcgagatatctgtatcttctgtcgccgacacgttacgtagattgtatggttc tgcggagtagtggacgagctgtacaagtaaccaggcatcaaataaaacgaaaggctcagtcgaaag actgggcctttcgttttatctgttgtttgtcggtgaacgctctctactagagtcacactggctcac cttcgggtgggcctttctgcgtttatactcgagtttattttgactaataatgacctacttaccatg ggtatggacagttttccctttgatatgtaacgcacgttgtggatatctggcgaaaatgagacgttg atcggcacgtaagaggttccaactttcaccataatgaaataagatcactaccgggcgtatttttg agttatcgagattttcaggagctaaggaagctaaaatggagaaaaaaatcactggatataccaccg

	ttgatatatcccaatggcatcgtaaagaacattttgaggcatttcagtcagttgctcaatgtacct ataaccagaccgttcagctggatattacggccttttaaagaccgtaaagaaaaataagcacaagt tttatccggcctttattcacattcttgcccgcctgatgaatgctcatccggaattccgtatggcaa tgaaagacggtgagctggtgatatgggatagtgttcacccttgttacaccgttttccatgagcaaa ctgaaacgttttcatcgctctggagtgaataccacgacgatttccggcagtttctacacatatatt cgcaagatgtggcgtgttacggtgaaaacctggcctatttccctaaagggtttattgagaatatgt ttttcgtctcagccaatccctgggtgagtttcaccagttttgatttaaacgtggccaatatggaca acttcttcgcccccgttttcaccatgggcaaatattatacgcaaggcgacaaggtgctgatgccgc tggcgattcaggttcatcatgccgtctgtgatggcttccatgtcggcagaatgcttaatgaattac aacagtactgcgatgagtggcagggcggggcgtaacctaggacgtcggctccttttggagcctttt tttttggcgcgccactgaccactcaactggaacaggcctgggagctagcgaaacagcgtttcgcgg cggtggggattgatgtcgaggaggcgctgcgccaacttgatcgtttac-3’

Table S4. Related to Figure 5.

Sequence of tc-acidTRP-09 ordered as gblock from IDT.

Primer	Sequence
TS1	atgagtaaaggagaagaacttttcactg
TS2	catcgaggtgaagacgaaagg
LS17	cgcttaatgctttccagatggcaac
LS18	aagtatgtggcctgacgcgtg
LS25	gcgcttcatgatgcgaatataagccctacggcggcctgccg
LS26	gacattcccggcaggccgccgtaggcttatattcgcatcatgaagcgc
ADN01	gataacatgcacatcatcgagatg
ADN02	ttacaaacattaataacgaagagatgacag
ADN03	ctcctgaaaatctcgataactcaaaaaatac
ADN04	gcttattaaaagcattctgtaacaaagc
TC36	gtacgtggtctcaggtcaacgctgtaatttattcagcgtttgtacatatc
TC37	gtacgtggtctcaggttttcatttaaccccactgcgg
TC38	gtacgtggtctcaaacctgctttttatactaagttggcattataaaaaagcattg
TC39	gtacgtggtctcatttcaagttagtataaaaaagctgaacgagaaacgtaaaatg
TC40	tacgtggtctcagaaacaaattgagggtatgacaatgagcaaaggagaagaacttttcac
TC47	gtacgtggtctcagacctcgttactgacaggaaaatgggcc

Table S5. Related to Methods. Full sequence of all oligos described in the methods.